

玉米自交系模糊聚类分析

曹靖生 曹大伟

(黑龙江省农科院作物育种所)

玉米是生产上利用杂种优势最早的农作物之一。人们在长期的育种实践中,积累了丰富的宝贵经验,如双亲互补、亲缘关系及地理位置较远等都是组配杂交种所需考虑的重要依据。但由于构成产量因素及其它农艺性状等因素的复杂性,仅仅根据双亲的表现型等进行亲本选配,往往不能得到满意的结果。近年来,我国许多育种工作者利用多元分析法测定与产量有关的数量性状的遗传距离,进行聚类分析或利用模糊数学对若干产量性状和与之有关的数量性状进行模糊聚类分析,以此作为一个综合指标,来预测作物的杂种优势,取得了一些较好的结果。

本试验对我省常用的玉米自交系进行了模糊聚类分析,拟探索现常用自交系在系统进化中所处的地位及彼此间的相互关系,以增强玉米杂交种组配的目的性,提高育种效率。

材料和方法

一、材料

本试验于1986年在黑龙江省农科院试验地进行,应用我省常用的40个玉米自交系。田间试验采用随机区组法,重复三次,双行区,4.5米行长,株行距30×70厘米。每品系调查8个性状:株高、穗位、穗长、穗粗、穗行数、行粒数、百粒重和每穗重。

二、统计分析方法

模糊聚类分析步骤:

1. 进行相似矩阵的标定。设 U 为被分类的对象全体,建立 U 上的相似关系 R 即求得相似矩阵。

2. 寻找模糊等价关系。相似关系 R 一般只满足反身性、对称性,采用二次幂方法,即 $R \rightarrow R^2 \rightarrow R^4 \dots$,经过有限步(步数 $\leq \lg N / \lg 2$)就可将 R 改造成为模糊等价矩阵。

3. 依等价模糊矩阵 R^* 对 U 进行分类。改造后矩阵中元素表示待分类对象彼此间相似的程度。规定任意的 λ ($0 \leq \lambda \leq 1$),使矩阵中元素

$$r_{ij} = \begin{cases} 1 & \text{若 } r_{ij} \geq \lambda \\ 0 & \text{否则} \end{cases}$$

则得到一个普通的等价关系,即决定了一个分类,选择适当的 λ 水平,即可得到相应的分类。亦按照 λ 水平,将样品间的关系用谱系图形式表示。

在本试验中,由于8个性状的物理单位不同,采用标准差标准化处理,即:

$$w_{ij} = \frac{(x_{ij} - \bar{x}_i)}{s_i} \quad (i = 1, 2, \dots, 40, \\ j = 1, 2, \dots, 8)$$

$$s_i = \sqrt{\frac{1}{n} \sum_{j=1}^n (w_{ij} - \bar{w}_i)^2}$$

将本试验数据采用上述公式进行标准化处理。

令 D_{ij} 表示论域中第 i 个元素(即第 i 组资料)与第 j 个元素的距离,根据:

$$D_{ij} = \sqrt{\sum_{k=1}^8 (w_{ik} - w_{jk})^2} \\ (\phi, j = 1, 2, \dots, 40)$$

求得 D_{ij} , 则 D_{ij} 的大小反映了二元素间的差异大小。这样便可以进行模糊聚类分析。

取定论域 $u = \{1, 2, 3, \dots, 39, 40\}$, 建立隶属函数(相似系数),

$$r_{ij} = 1 - D/M$$

M 为一个适当选定的常数, 使 r_{ij} 在 0 与 1 之间取值。由 r_{ij} 与 D_{ij} 的关系知, 当 r_{ij} 愈大(小), 第 i 元素与第 j 元素之间的差异愈小(大)。

再取 R 的乘幂, 将 R 变成为:

$$R^{(k)} = R^0 = R^* = (r_{ij})$$

令 $R_\lambda = (\lambda_{ij})$, 则 R_λ 称为 λ 的图象。

结果与分析

按上述步骤进行分析, 得到 R_λ , 即 λ_{ij} , 见表。 R_λ 是普通的等价关系, 它给出了 u 的一分类方法。我们对不同的 λ 值进行计算, 得到如下分类结果:

(1) 取 $\lambda = 1$, 根据 R_λ , 40 个自交系各自为一类;

(2) 取 $\lambda = 0.94$, 由 R_λ 将自交系 4、15 归为一类, 其余各自为一类;

(3) 取 $\lambda = 0.93$, 聚为 $[4, 15]$, $[8, 28]$, 余者各为一类;

(4) 取 $\lambda = 0.90$, 聚为 $[4, 15]$, $[8, 28]$, $[1, 2, 17, 19, 40]$, 余者各自为一类;

(5) 取 $\lambda = 0.89$, 聚为 $[4, 5]$, $[8, 28]$, $[1, 2, 17, 19, 40]$, $[27, 30]$, 其余各为一类;

(6) 取 $\lambda = 0.88$, 聚为 $[4, 15]$, $[1, 2, 8, 9, 17, 19, 21, 23, 24, 26, 27, 28, 30, 40]$, 余者各为一类;

(7) 聚 $\lambda = 0.87$, 聚为 $[4, 15]$, $[1, 2, 5, 6, 8, 9, 17, 18, 19, 21, 22, 23, 24, 26, 27, 28, 30, 32, 40]$, 其余各为一类;

(8) 取 $\lambda = 0.86$, 聚为 $[1, 2, 4, 5, 6, 8, 9, 11, 17, 18, 19, 21, 22, 23, 24, 26, 27, 28, 32, 40]$, $[3, 38]$, 余者各为一类;

(9) 取 $\lambda = 0.85$, 聚为 $[1, 2, 4, 5, 6, 7, 8,$

$9, 11, 17, 18, 19, 21, 22, 23, 24, 26, 27, 28, 32, 33, 40]$, $[3, 14, 38]$, 余者各为一类;

(10) 取 $\lambda = 0.84$, 聚为 $[1, 2, 4, 5, 6, 7, 8, 9, 11, 16, 18, 19, 21, 22, 23, 24, 26, 27, 28, 32, 33, 40]$, $[3, 14, 38]$, $[29, 35]$, 其余各为一类;

(11) 取 $\lambda = 0.83$, 聚为 $[1, 2, 4, 5, 6, 7, 8, 9, 11, 16, 19, 21, 22, 23, 24, 26, 27, 28, 29, 32, 33, 35, 40]$, $[3, 14, 38]$, 余者各为一类;

(12) 取 $\lambda = 0.82$, 聚为 $[1, 2, 4, \dots, 9, 11, 16, 18, 19, 21, 22, 23, 24, 26, 27, 28, 29, 32, 33, 35, 39, 40]$, $[3, 14, 38]$, 余者各为一类;

(13) 取 $\lambda = 0.81$, 聚为 $[1, 2 \dots 9, 21 \dots 24, 26, 27 \dots 31, 32 \dots 36, 38, 39, 40]$, $[20]$, $[25]$, $[31]$, $[36]$ 五类;

(14) 取 $\lambda = 0.80$, 聚为 $[1, 2, \dots, 19, 21 \dots 24, 26 \dots 40]$, $[20]$, $[25]$ 三类;

(15) 取 $\lambda = 0.74$, 聚为 $[1, 2 \dots 24, 26 \dots 40]$, $[25]$ 二类;

(16) 取 $\lambda = 0.58$, 全部自交系聚为一类。

将上述结果作成模糊聚类谱系图(图1)。根据上述模糊聚类分析, 我们可以得到下面几点结论:

1. 在 40 个自交系中, 有些自交系是由某个自交系转育了个别或一些性状后选育来的。模糊聚类分析也表明, 它们之间的分类系统往往较近。如 4(吉 63^H)与 15(吉 63)等;

2. 由某个自交系同另一个自交系的杂种后代选育而来的自交系, 可以形成同亲本分类系统较远的不同分类系统。如 458 瓦 742A 虽是由 458 × 瓦 742A 的杂种后代选育而来的, 但 458 与 458 瓦 742A 之间的分类系统却较远;

3. 具有相同亲本之一的二环系选育而来的自交系, 它们之间的分类系统也并非很近, 如 13 (153 M017)、19(小金153)、23(153A634)三个自交系中, 小金153与153A634系之间分类系统较近, 而这二个自交系与153M017之间的分类系统却较远, 可能是

153M017 中 M017 所占遗传比重较大之故。

4. 自交系亲缘的远近和彼此间分类差别, 可以作为组杂交种的重要依据。如在目前生产上和试验中表现较好的大风 71×OH43^{HT}(龙单三号)、大风71×28RO(龙210)、吉 63^{HT}×M14^{HT}(龙 127)、吉 818×系 14(龙 124)等, 这些自交系二二之间的分类系统都较远, 因而表现出的杂种优势较大。至于同一类型中具体选择哪个自交系作为亲本, 还要根据成熟期、抗病性、抗逆性等具体条

件, 结合育种目标来定, 遵循杂交亲本选配的一般原则, 这样既能提高选配的效果, 又能减少盲目选配的数量, 提高育种效率。

结 语

1. 通过模糊聚类分析, 不仅仅限于数量分类, 还可以为二环系的选育、改良群体的组成提供科学依据。因为不同类群间杂交可使性状得到互补, 又免使新选育系的遗传基

表 1 自交系编号表

类 号	自交系名称	类 号	自交系名称	类 号	自交系名称	类 号	自交系名称
1	黄早 4	11	吉 830	21	RW64A	31	冬 96
2	OH43 ^{HT}	12	458	22	153A634	32	64O103
3	红玉米	13	153M017	23	413	33	L105
4	吉63 ^{HT}	14	73-1	24	大四	34	MA21
5	28RO	15	吉63	25	桦94	35	系14
6	赤401	16	大 黄46	26	东46	36	原502
7	英64-1	17	海014	27	长 8	37	903
8	冬 黄	18	小金153	28	吉815	38	安441
9	B73 ^{HT}	19	吉818	29	大风71	39	458瓦742A
10	O103长3	20	M14 ^{HT}	30	458RP ₁	40	罗31

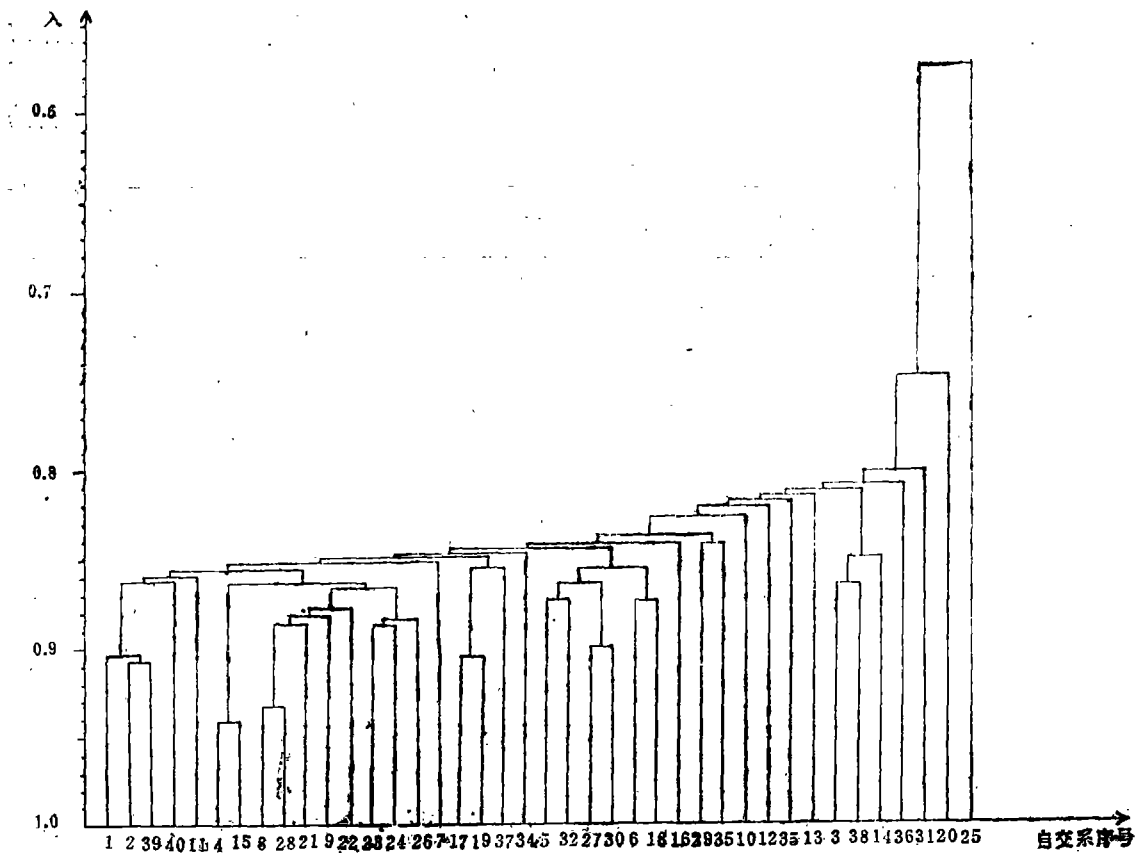
表 2 不同 λ 水平下的分类情况

序 号		类 号		λ 值	序 号		类 号		λ 值
1	4	15	0.941952812		21	17	37	0.855930747	
2	8	28	0.931514717		22	1	7	0.855877019	
3	2	39	0.907757272		23	8	14	0.853813527	
4	17	19	0.904521844		24	1	17	0.853295412	
5	1	2	0.903172874		25	1	33	0.852136321	
6	27	30	0.899578314		26	5	6	0.850692993	
7	23	24	0.888361405		27	1	5	0.849902464	
8	23	26	0.888099773		28	1	16	0.845208258	
9	8	21	0.885449342		29	29	35	0.844264355	
10	8	9	0.880831421		30	1	29	0.837547793	
11	8	22	0.879976560		31	1	10	0.824640543	
12	5	32	0.871160849		32	1	12	0.819716826	
13	6	18	0.870275484		33	1	34	0.814018525	
14	8	23	0.867095712		34	1	13	0.813232508	
15	4	8	0.865987677		35	1	3	0.811697092	
16	1	40	0.863442383		36	1	36	0.809820204	
17	1	11	0.863423377		37	1	31	0.801034116	
18	5	27	0.862534279		38	1	20	0.748863810	
19	3	38	0.862118347		39	1	25	0.584887802	
20	1	4	0.858340198						

础日趋狭窄，给选配带来困难。选择不同类型的自交系组成群体可使群体的基因型丰富，提高轮回选择的效果。

2. 利用模糊数学这个数学工具，对现有自交系进行数量分类，对于我们还仅是一个

尝试，许多问题还有待于进一步探讨如哪些性状作为分类的依据可靠性最好、理论上尽管性状愈多，其分类结果愈可靠，但势必又造成工作量和计算量的增加等问题。如何解决这些问题尚须进一步研究。



40个自交系模糊聚类谱系图

3. 模糊聚类分析应用于玉米自交系的数学分类，固然有助于提高我们的分类水平，减少主观盲目性，但同其它数学方法一样，它并非是万能的，因此，在使用过程中，应联系我们的实际情况，正确地使用这个教学工具。

