

文章编号: 1005-0906(2006)02-0016-03

玉米引种材料的主成分分析和聚类分析

刘玉爱, 侯建华, 高志军, 周 伟

(内蒙古农业大学农学院, 呼和浩特 010019)

摘 要: 对 24 个热带玉米引种材料的株高、穗位高、穗长、穗粗、穗行数、行粒数、百粒重、穗上位叶面积、穗位叶面积、穗下位叶面积、抽雄期、粗蛋白和粗脂肪等 20 个性状进行了测定, 对得到的各性状值进行主成分分析。结果表明: 形态和产量因子、生育期因子、穗行数因子、低产因子、脂肪含量因子、蛋白质含量因子和穗长因子这 7 个主要因子对变异的累积贡献率达 83.149%。依据主成分向量将 24 个热带玉米引种材料聚为 6 类。

关键词: 玉米; 种质资源; 主成分分析; 聚类分析

中图分类号: S513.024

文献标识码: A

Principal Component Analysis and Cluster Analysis of Maize Introduced Varieties

LIU Yu-ai, HOU Jian-hua, GAO Zhi-jun, ZHOU Wei

(College of Agronomy, Inner Mongolia Agriculture University, Huhehaote 010019, China)

Abstract: It was tested to plant height, ear height, ear length, ear diameter, row number per ear, kernel number per row, weight of 100 grain, leaf area above the ear, leaf area of the ear, leaf area under the ear, head sprouting, crude protein and oil which 20 traits of 24 tropical maize introduced varieties were involved were tested and component analysis was performed for them. The results indicated that the first seven principal components which accounts for over 83.149% of total variations were: morphology and yield factor, growth duration factor, row number per ear factor, low yield factor, oil content factor, protein content factor and ear length factor. Using them as complex index, the 24 tropical maize introduced varieties were grouped into 6 clusters.

Key words: Maize; Germplasm resources; Principal component analysis; Cluster analysis

一个新品种的育成, 离不开优异种质资源的利用。在有限的种质资源内, 新品种的选育潜力会很小^[1]。热带亚热带玉米种质是在自然和人为选择的共同作用下逐渐演变形成的, 在植株、果穗及子粒性状方面具有丰富的遗传多样性。长期以来与其它种质类群的遗传交流较少, 遗传差异较大, 形成了独特的种质类群, 是温带玉米育种不可多得的异源种质。利用先进的手段和技术把热带亚热带玉米种质导入温带种质, 进行种质扩增、改良和创新研究, 具有重要的现实意义^[2]。为此, 我们以引入的 24 个具有热带玉米血缘的育种原始材料为试材, 进行了田间及室

内各形态、产量、品质等性状的观测, 并对取得的各性状值进行了主成分和聚类分析。以此对引种材料进行综合评价, 为进一步选育优良自交系提供选择依据。

1 材料与方法

供试材料为 24 个具有热带玉米血缘的引种材料(引自中国农科院作物所), 分别为 S-266、S-267、S-268、S-269、S-270、S-271、S-272、S-273、S-274、S-275、S-276、S-277、S-279、S-280、S-281、S-282、S-283、S-284、S-285、S-297、S-298、S-310、S-321、S-322, 2003 年 4 月 25 日种植于内蒙古农业大学教学农场。田间试验采用完全随机区组设计, 3 次重复, 双行区种植, 行距为 65 cm, 株距为 25 cm, 行长为 6 m, 各品系随机取 10 株套袋自交, 抽雄后调查株高、穗位高、雄穗分枝数、茎秆粗度、叶片数、穗上位叶面积、穗位叶面积、穗下位叶面积、抽雄期; 单

收稿日期: 2005-04-14

基金项目: 内蒙古自然科学基金项目(200208020315)

作者简介: 刘玉爱(1978-), 女, 硕士研究生, 从事玉米遗传育种

Tel: 0471-4318396 E-mail: liuyuai612@163.com

侯建华为本文通讯作者。Tel: 0471-4308656

E-mail: houjh68@163.com

株收获后风干于室内考种,测得穗长、穗粗、秃尖长、穗行数、行粒数、穗重、轴重、百粒重、穗粒重 9 个性状。每个品系随机取 3 个平均样,用半微量凯氏定氮法测子粒的粗蛋白值,用索氏提取法检测子粒的粗脂肪值,共求得 20 个性状的平均值。

用 SPSS 软件^[3]对以上数据进行方差分析、主成分分析和聚类分析。

2 结果与分析

2.1 方差分析

对 20 个性状进行方差分析结果表明,供试材料各性状之间差异均达显著水平,说明这 20 个性状在 24 个材料间存在真实差异,可做进一步分析。

2.2 主成分分析

对 20 个性状进行主成分分析的结果表明(表 1),前 7 个主成分的累积贡献率达 83.149%,因此,可用前 7 个主成分概括这些性状的总信息量,并以此对这些品系进行综合评价。

表 1 主成分分析的特征向量系数

项 目	特 征 根						
	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
	4.307	3.404	2.686	2.331	1.488	1.311	1.102
株 高	0.587	-0.273	-0.499	-0.014	-0.344	0.048	0.226
穗位高	0.634	0.021	-0.519	-0.021	0.031	0.482	-0.014
雄穗分枝数	0.018	0.307	-0.823	-0.023	-0.057	-0.284	0.114
茎 粗	-0.405	0.475	0.214	0.235	-0.318	0.194	0.372
叶片数	-0.242	0.412	-0.321	0.471	0.120	-0.166	0.059
穗上位叶面积	0.794	-0.073	0.212	0.314	-0.199	-0.030	0.049
穗位叶叶面积	0.789	-0.049	0.300	0.440	-0.107	-0.118	0.077
穗下位叶面积	0.678	-0.113	0.075	0.551	0.034	-0.213	-0.083
抽雄期	-0.423	0.688	-0.164	-0.104	-0.105	-0.335	-0.071
穗 长	-0.055	0.024	0.252	-0.022	0.570	0.137	0.726
穗 粗	-0.181	0.346	-0.267	0.176	0.297	0.481	-0.330
秃尖长	-0.389	0.014	0.211	0.716	0.152	-0.115	-0.374
穗行数	0.117	-0.543	0.574	-0.198	0.214	-0.021	-0.089
行粒数	0.420	0.677	-0.028	-0.465	-0.088	0.054	-0.063
穗 重	0.558	0.647	0.338	-0.284	0.158	0.028	-0.126
轴 重	0.015	0.680	0.400	0.140	0.233	-0.138	0.124
百粒重	0.163	0.393	-0.058	0.635	-0.042	0.457	0.060
穗粒重	0.682	0.473	0.245	-0.365	0.093	0.041	-0.201
粗蛋白	-0.506	-0.028	0.393	-0.098	-0.397	0.500	-0.080
粗脂肪	0.099	-0.274	-0.416	-0.031	0.670	0.108	-0.082
贡献率(%)	21.537	17.018	13.432	11.657	7.439	6.555	5.512
累计贡献率(%)	21.537	38.554	51.987	63.643	71.082	77.637	83.149

由表 1 可知,第一主成分的特征根 $\lambda_1=4.307$,对总遗传贡献率最大,为 21.537%。对应特征向量中,数值较大的性状依次是穗上位叶面积、穗位叶叶面积、穗粒重、穗下位叶面积、穗位高、株高、穗重、行粒数、百粒重和穗行数。这一特征向量中,穗叶面积和穗粒重最大,其中粗蛋白含量、茎粗、穗粗的负向系数大。说明叶片宽大、植株高大的品系产量高,但粗蛋白含量低,第一主成分可作为形态和产量因子。第二主成分的特征根 $\lambda_2=3.404$,对总遗传方差贡献率为 17.018%。对应特征向量中,以抽雄期、轴重、行粒数、穗重、茎粗、穗粒重、叶片数、百粒重的值较大,说明生育期与产量存在正相关关系,第二主成分可作为生育期因子。由于第一、第二主成分都反映了产量性状,所以把第一、第二主成分归为高产因子。第三主成分的特征根 $\lambda_3=2.686$,对总遗传贡献率达 13.432%。对应特征向量中,穗行数、轴重、粗蛋白含量、穗重的值较大,第三主成分可作为穗行数因子。

第四主成分的特征根 $\lambda_4=2.331$,对总遗传贡献率为 11.657%。在第四主成分的特征向量中,秃尖长、百粒重的值最大,而穗重、穗粒重的值为负值。说明秃尖越长,单穗产量越低,第四主成分可定为低产因子。第五主成分的特征向量中,粗脂肪含量、穗长的值较大,而粗蛋白、株高、穗叶面积为负值,第五主成分可作为脂肪含量因子。要选育高脂肪品系,可间接选择长穗型。第六主成分的特征向量中,粗蛋白含量、穗位高、穗粗、百粒重的值依次较高,可作为蛋白含量因子。若选育高蛋白材料,可以选穗位高和穗粗值较大的品系。第七主成分的特征向量中,穗长的值明显大于其它值,因此,第七主成分主要反映了穗长因子。

由入选的特征值和特征向量计算出 24 个材料的各综合指标的 7 个主成分值: g_1 、 g_2 、 g_3 、 g_4 、 g_5 、 g_6 和 g_7 ,并以此作为选择的参考依据,进行优良材料的选择。第一主成分值越大,则植株高大,叶面积宽

大,也有利于产量的提高。这类材料适应性强,高产稳产,但受蛋白含量、秃尖长、穗粗的抑制,因此,蛋白含量不易太高,秃尖不宜太长,穗粗适当为好。以粗蛋白因子考虑,第六主成分值较大为好。以粗脂肪因子考虑,第五主成分值大较好。以穗长因子考虑,第七主成分值越大越好。若要增加穗行数,则第三主成分值越大越好。从提高产量角度考虑,第一、二主成分值越大越好,第四主成分值越小越好。

2.3 聚类分析

供试材料的每个主成分值可组成一个主成分向量,两个材料间的主成分向量距离平方和为两材料间有关性状的遗传距离。依据主成分向量用最短距离法计算材料间的欧氏距离,将供试的 24 个材料分为 6 类(图 1),并根据主成分值和原始数据对聚类的结果进行分析。第一类材料为 S-267、S-271、S-277、S-270、S-310、S-269、S-280、S-275,其特点为植株高大,叶面积大,产量高,但生育期较长,蛋白含量和脂肪含量低。第二类为 S-297、S-298、S-284、S-285、S-283、S-322、S-273、S-281、S-272、S-268、S-282、S-279,农艺性状表现次于第一类,而品质比第一类好,属于中等型。第三类为 S-276,属早熟品系,植株高度中等,果穗偏长,产量中等偏低,蛋白含量较高。第四类为 S-266,其特点是穗行数少,秃尖长,百粒重高,产量低,脂肪含量高。第五类材料为 S-310,它的总体农艺性状较差,品质一般,没有利用价值。第六类材料虽然植株高大,果穗长而且产量很高,但生育期太长,不易成熟。

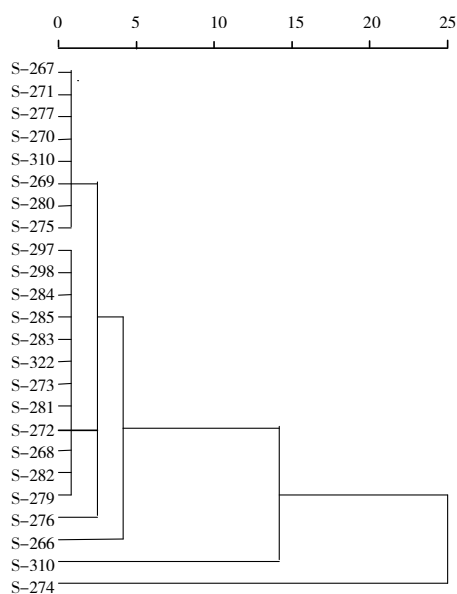


图 1 24 个材料的欧氏聚类图

3 结论与讨论

(1)通过主成分分析得出了形态和产量因子、生育期因子、穗行数因子、低产因子、脂肪含量因子、蛋白含量因子和穗长因子为玉米的 7 个主成分。每个主成分都比较客观地反映了所控制的各性状之间的相互关系。这 7 个主成分对变异的累计贡献率达 83.149%。从综合性状来看,这些材料的产量指标和品质指标不易同时达到优良水平,所以对它们的利用还需要进行局部改良,或者进行个别性状的利用,如采用杂交或回交将其优良性状导入其它玉米自交系中。

(2)在各种问题的多指标分析中,指标的个数多,又有相互关联,分析工作比较困难。多元统计的主成分分析是指在不损失或很少损失原有信息的前提下,将原来个数较多而且彼此相关的指标转换为新的个数较少而彼此独立或不相关的综合指标,从而简化多指标分析^[4]。很多研究表明,应用主成分分析方法进行综合评价是一种有效的方法^[5~9],但本文研究结果表明,主成分分析往往会提供相互矛盾的评价信息。例如,在第五主成分中,粗脂肪含量会随着穗长的加大而增高,而第七主成分则显示出两者呈负相关。所以用主成分分析评价一个材料优劣时,还需要结合原始数据,采用辩证的思维及分析方法,去除表面现象。

(3)主成分分析基础上的聚类,可有效地剔除一些无关大局的因子,使结果更加精确,在种质资源分析及评价中应用效果较好。

参考文献:

- [1] 金 镛,徐艳霞,李芳志,等. 从丹 408 的试种成功讨论玉米种质资源的利用[J]. 辽宁农业科学, 1998, (5): 36-36.
- [2] 刘治先. 热带亚热带玉米种质的改良研究进展[J]. 玉米科学, 2000, 8(3): 28-32.
- [3] 张文彤. SPSS11 统计分析教程[M]. 北京: 希望电子出版社, 2002. 190-202.
- [4] 袁志发,周静宇. 多元统计分析[M]. 北京: 科学出版社, 2003. 188-201.
- [5] 印志同,邓德祥,卞云龙,等. 玉米自交系性状的遗传相关分析和主成分分析[J]. 扬州大学学报, 2001, 4(1): 49-51.
- [6] 谭远德,鲁 成,向仲怀. 品种综合评价的数学方法[J]. 生物数学学报, 2000, 15(4): 457-466.
- [7] 魏亦农,曹连莆. 二棱啤酒大麦品种资源农艺性状的聚类分析和主成分分析[J]. 种子, 2003, (3): 69-70.
- [8] 傅同良. 33 个糯玉米自交系遗传主成分和聚类分析[J]. 中国农业科学, 1995, 28(5): 46-53.
- [9] 徐尚忠,宋德凤,吴高岭. 数值分类法在玉米品种资源上的应用[J]. 玉米科学, 2000, 8(1): 23-27.