

# 基于主成分分析和聚类分析的 玉米品种的综合评价

张中伟 杨海龙 付俊 谢文锦 李方明 高旭东

(丹东农业科学院 辽宁凤城 118109)

**摘要:**为综合评价东华北春玉米组的 33 个参试玉米品种,选取 14 个主要农艺性状进行系统比较分析,应用主成分分析和聚类分析综合评价并筛选优质玉米品种。结果表明,倒伏(折)率、空秆率、秃尖长的变异系数较大,生育期的变异系数最小;通过选取前 5 个主成分构建综合评价模型,最后应用聚类分析进行验证,将 33 个玉米品种分为 3 类,筛选出金园 23 和吉农玉 1898 两个优质玉米品种。

**关键词:**玉米;农艺性状;主成分分析;聚类分析

玉米是全世界重要的粮食作物之一,同时也是我国旱田作物中种植面积最大的作物,2020 年我国玉米种植面积达 6.2 亿亩,产量约 2 600 亿 kg<sup>[1]</sup>。玉米不仅是主要的粮食作物,又是重要的工业原料和饲料来源,东华北春玉米区是我国的玉米主产区之一和重要的商品粮基地,在国民经济中具有举足轻重的战略地位<sup>[2]</sup>。传统育种中,仅通过个别农艺性状表现和产量高低判断玉米品种的优劣具有局限性,对玉米品种多个性状指标进行全面分析和综合评价才更具有可靠性和客观性<sup>[3]</sup>。主成分分析是采取降维的方法,把多个指标转化为少数几个综合指标的一种统计方法,使这些综合指标尽可能反映原始数据的信息,从而对品种进行科学评价。聚类分析是研究事物分类的基本方法,在主成分基础上进行聚类分析可以有效去除一些影响较小的因子,便于对品种进行综合分析和全面评判<sup>[4]</sup>。笔者以东华北春玉米组的 33 个参试玉米品种为研究对象,选取 14 个主要农艺性状数据,综合应用主成分分析和聚类分析来评价品种的优劣,旨在为玉米新品种选育、审定和推广提供参考。

## 1 材料与方法

### 1.1 供试材料及试验设计

试验材料为国家区域试验东华北中晚熟春玉米

组各单位提供的 33 个玉米品种,其中以郑单 958 为对照品种,各品种编号和品种名称见表 1。试验于 2021 年在丹东农业科学院试验基地进行。试验采取随机区组排列,3 次重复,5 行区,小区面积 20 m<sup>2</sup>,试验密度为 67 500 株/hm<sup>2</sup>,四周设不少于 4 行的保护行。田间管理同当地大田。

### 1.2 调查项目

调查生育期( $X_1$ )、株高( $X_2$ )、穗位高( $X_3$ )、穗长( $X_4$ )、穗行数( $X_5$ )、秃尖长( $X_6$ )、穗粗( $X_7$ )、轴粗( $X_8$ )、行粒数( $X_9$ )、百粒质量( $X_{10}$ )、出籽率( $X_{11}$ )、空秆率( $X_{12}$ )、倒伏(折)率( $X_{13}$ )和产量( $X_{14}$ )14 个性状。每次重复实收中间 3 行风干脱粒,产量按 14%含水量计算。其他农艺性状均按国家区试标准观察、考种和记载。

### 1.3 数据分析

采用 Excel 整理各品种的产量数据和主要农艺性状数据。用 SPSS 26.0 进行主成分分析和聚类分析。

## 2 结果与分析

### 2.1 玉米品种主要农艺性状变异分析

由表 2 可知,参试玉米品种的 14 个农艺性状的变异各不相同,变异系数幅度为 1.08%~148.37%,倒伏(折)率、空秆率、秃尖长的变异系数较大,分别为 148.37%、101.79%和 68.98%。生育期、出籽率的变异系数较小,分别为 1.08%和 2.26%。说明倒伏(折)率

作者简介:张中伟(1987-),女,硕士,助理研究员,从事玉米品种试验工作。E-mail:zhangzhongwei03@163.com

表 1 参试玉米品种名称及编号

编号	品种名称	编号	品种名称	编号	品种名称
1	先玉 2056	12	宏育 107	23	NY101
2	吉农大 2088	13	雅玉 6910	24	金园 59
3	LY1210	14	鲁单 510	25	辽单 805
4	金园 23	15	新丹 166	26	佳昌 999
5	新玉 158	16	HT2101	27	润民 966
6	禾育 207	17	京科 4515	28	QS519
7	L2128	18	宏硕 2001	29	万玉 156
8	HA502	19	吉农玉 1898	30	LY2118
9	强盛 363	20	强盛 379	31	辽科 666
10	京科 8568	21	利合 1088	32	先玉 2118
11	承玉 100	22	L2016	33	郑单 958(CK)

和空秆率受自然条件和品种影响较大,在玉米品种选育时,抗倒伏(折)性状的选择范围更大,说明本地区适宜筛选出抗倒性好的玉米品种。

2.2 不同玉米品种主要农艺性状的主成分分析

对参试玉米品种的 14 个农艺性状进行主成分分析,首先对原始数据进行标准化处理,然后进行 KMO 检验,得到 KMO 值=0.535>0.500;对其进行 Bartlett 球形度检验,得到卡方值=247.993, Sig.<0.05,进一步说明各性状间存在较强的相关性,适宜进行主成分分析。

根据特征值大于 1 的标准,选取前 5 个主成分,其累积贡献率达 77.058%(表 3),包含了 14 个农艺性状的绝大部分信息,因此,可以选择这 5 个主成分作为玉米品种的综合评价指标。

主成分特征向量值反映了各农艺性状在主成分中的影响力大小和方向,由表 3 可知,主成分 1 主要反映倒伏(折)率、空秆率、穗粗和轴粗性状;主成分 2 主要反映穗长、行粒数和产量性状;主成分 3 主要反映株高性状;主成分 4 主要反映穗位高性状;主成分 5 主要反映百粒重性状。主成分分析结果表明抗倒性指标已经成为评判一个玉米品种优劣的重要性状,玉米育种中应综合选择产量和抗性指标,从而筛选出优质、高产、抗性优良的玉米品种。

2.3 不同玉米品种的综合评价

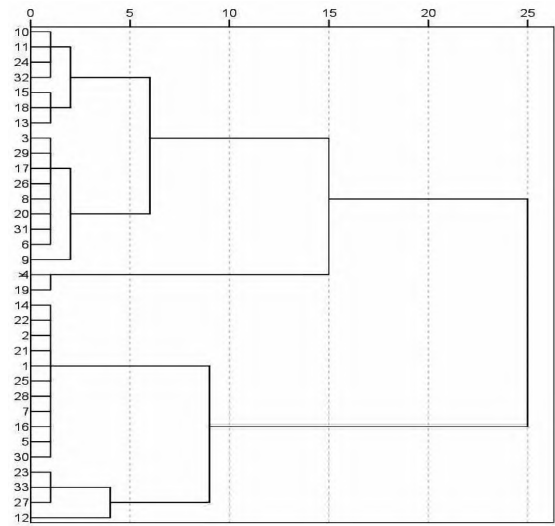
根据表 3 中各主成分的特征值,计算 33 个玉米品种的主成分因子得分,再以每个主成分的贡献率为权重,构建出不同玉米品种的综合评价模型: $F=$

$0.266Y_1+0.185Y_2+0.133Y_3+0.107Y_4+0.080Y_5$ (表 4)。

由表 4 可知,在 33 个玉米品种中,综合得分较高的品种为金园 23 和吉农玉 1898,这 2 个玉米品种在产量和抗性等综合性状上表现都比较好。

2.4 不同玉米品种综合得分的聚类分析

利用 SPSS 26.0 软件,将各品种对应综合得分采用系统聚类中平方 Euclidean 距离聚类和平均联接(组间)构建系统聚类图(附图),在遗传距离为 10 处将 33 个玉米品种分为 3 大类。第 1 类是综合性状较优的品种,共有 2 个,分别为金园 23 和吉农玉 1898,这一类群品种综合性状较好、产量较高;第 2 类群为综合性状居中的品种,有 16 个品种;第 3 类群的品种有 15 个,这类品种综合评分较低。



附图 28 个玉米品种 14 个农艺性状的系统聚类分析

表 2 参试玉米品种主要农艺性状及变异分析

编号	性状													
	X <sub>1</sub> (d)	X <sub>2</sub> (cm)	X <sub>3</sub> (cm)	X <sub>4</sub> (cm)	X <sub>5</sub> (行)	X <sub>6</sub> (cm)	X <sub>7</sub> (cm)	X <sub>8</sub> (cm)	X <sub>9</sub> (粒)	X <sub>10</sub> (g)	X <sub>11</sub> (%)	X <sub>12</sub> (%)	X <sub>13</sub> (%)	X <sub>14</sub> (kg/hm <sup>2</sup> )
1	129	315	141	20.5	16.8	0.4	5.0	2.8	34.4	38.3	86.7	6.2	4.9	10 519.5
2	130	302	123	18.7	15.6	0.9	5.1	3.0	33.2	39.3	84.4	4.5	0	11 635.5
3	128	304	141	19.4	15.2	0.1	5.3	3.0	31.8	45.8	81.1	3.7	5.5	11 706.0
4	128	338	149	21.2	15.6	1.0	5.3	3.3	34.4	49.9	83.8	6.2	5.5	11 794.5
5	131	286	136	18.5	17.6	0.1	5.4	3.1	38.2	37.2	85.4	1.3	1.3	12 156.0
6	130	321	133	20.4	17.2	0.8	5.7	3.1	38.2	43.8	86.4	5.9	8.8	11 259.0
7	127	317	116	21.1	16.8	0.5	5.0	3.1	38.6	39.7	88.6	0.8	1.8	12 081.0
8	131	287	140	19.8	17.6	1.3	5.4	3.1	37.0	43.1	84.5	19.6	15.2	9 378.0
9	131	314	110	19.9	17.2	0.5	5.5	3.3	35.4	45.8	82.8	20.2	18.9	9 058.5
10	129	337	150	19.3	17.6	0.2	5.3	3.2	37.0	39.9	86.2	0.4	0	12 835.5
11	131	292	126	20.2	18.4	0.7	5.6	3.2	37.4	41.9	83.7	8.7	0	11 142.0
12	130	303	115	15.9	15.6	0.5	5.0	2.7	30.2	40.2	87.7	2.9	1.3	11 256.0
13	129	316	138	21.2	14.8	1.4	5.1	2.8	38.8	41.3	86.3	2.5	0	11 902.5
14	125	279	112	20.2	17.6	0.8	5.3	3.2	37.0	37.5	86.3	1.2	0	11 856.0
15	129	284	123	19.3	17.6	0.5	5.3	3.0	38.2	40.0	83.9	11.3	8.8	9 580.5
16	130	292	132	18.9	16.0	0.1	5.2	2.9	34.6	43.8	86.0	0.8	0	12 274.5
17	131	302	137	20.9	16.0	0.8	5.4	3.2	39.2	45.1	85.8	7.5	2.5	11 397.0
18	131	279	118	20.6	16.4	1.8	5.3	3.1	36.8	43.5	85.7	1.7	0	12 766.5
19	131	318	126	21.9	19.2	1.5	5.7	3.5	37.4	45.6	82.5	4.2	6.8	11 719.5
20	131	329	132	19.0	16.8	0.9	5.4	3.1	34.8	45.1	85.1	17.0	14.0	8 997.0
21	130	292	130	18.9	14.4	0.2	5.1	3.0	37.4	38.1	85.1	6.6	0	11 458.5
22	130	273	121	19.6	15.6	0.1	5.3	2.9	34.8	46.6	89.5	3.3	0	11 875.5
23	130	273	121	18.3	18.0	0	5.1	3.0	32.8	37.8	84.1	0.8	0	11 994.0
24	129	292	129	20.2	18.0	1.0	5.4	3.1	39.4	42.1	84.1	7.5	0	10 566.0
25	130	292	141	17.0	18.0	0.4	5.5	3.3	34.6	37.1	85.5	2.9	1.7	11 833.5
26	131	249	119	20.2	17.6	1.0	5.9	3.5	37.0	43.5	80.7	3.3	4.3	11 866.5
27	131	268	125	15.2	18.4	1.4	5.3	3.1	24.6	39.1	85.2	14.5	0	9 189.0
28	129	302	122	19.1	16.8	1.1	5.4	3.1	36.0	39.9	85.6	3.8	0	11 878.5
29	129	305	132	20.7	19.2	0.6	5.5	3.2	37.0	42.9	83.9	2.5	2.8	12 175.5
30	130	298	136	19.1	17.2	0.1	5.1	2.9	30.6	40.7	82.6	0.4	3.2	12 078.0
31	128	317	138	19.9	17.6	0.7	5.6	3.2	36.4	45.2	83.7	0.8	1.5	12 228.0
32	128	308	132	21.0	16.0	0.7	5.3	3.0	40.8	44.1	87.1	4.1	0	11 686.5
33	131	273	138	16.4	16.6	0.3	5.3	3.2	31.1	37.9	85.7	0	0	10 958.0
平均	129.6	292.4	129.8	19.5	16.9	0.7	5.3	3.1	35.6	41.9	85.0	5.4	3.3	11 366.7
标准差	1.41	36.07	10.23	1.55	1.19	0.47	0.21	0.18	3.29	3.27	1.92	5.46	4.89	1 044.67
变异系数(%)	1.08	12.34	7.89	7.96	7.04	68.98	4.03	5.79	9.23	7.80	2.26	101.79	148.37	9.19

注: X<sub>1</sub> 为生育期; X<sub>2</sub> 为株高; X<sub>3</sub> 为穗位; X<sub>4</sub> 为穗长; X<sub>5</sub> 为穗行数; X<sub>6</sub> 为秃尖长; X<sub>7</sub> 为穗粗; X<sub>8</sub> 为轴粗; X<sub>9</sub> 为行粒数; X<sub>10</sub> 为百粒质量; X<sub>11</sub> 为出籽率; X<sub>12</sub> 为空秆率; X<sub>13</sub> 为倒伏(折)率; X<sub>14</sub> 为产量。下同。

表 3 参试玉米品种的 14 个农艺性状的主成分载荷矩阵和特征值

性状	主成分				
	1	2	3	4	5
X <sub>1</sub>	0.365	-0.464	-0.312	-0.085	0.454
X <sub>2</sub>	0.182	0.126	0.492	0.661	-0.303
X <sub>3</sub>	-0.021	0.189	0.278	0.667	0.379
X <sub>4</sub>	0.325	0.759	0.388	-0.231	-0.044
X <sub>5</sub>	0.474	0.051	-0.563	0.218	-0.459
X <sub>6</sub>	0.469	0.109	-0.042	-0.465	-0.025
X <sub>7</sub>	0.754	0.317	-0.404	0.032	0.091
X <sub>8</sub>	0.680	0.364	-0.438	0.139	-0.054
X <sub>9</sub>	0.208	0.701	0.268	-0.243	-0.253
X <sub>10</sub>	0.519	0.292	0.357	-0.179	0.538
X <sub>11</sub>	-0.589	-0.049	0.237	-0.397	-0.196
X <sub>12</sub>	0.688	-0.558	0.346	-0.140	-0.110
X <sub>13</sub>	0.748	-0.292	0.408	0.046	-0.052
X <sub>14</sub>	-0.530	0.719	-0.280	0.061	0.225
特征值	3.722	2.583	1.859	1.501	1.124
贡献率(%)	26.585	18.450	13.278	10.719	8.026
累积贡献率(%)	26.585	45.035	58.313	69.032	77.058

### 3 讨论与结论

产量相关农艺性状变异系数的大小反映了玉米品种丰产性和稳产性的表现<sup>[5]</sup>。本研究 33 个玉米品种的 14 个主要农艺性状变异系数为 1.08%~148.37%,其中倒伏(折)率、空秆率、秃尖长的变异系数较大且均在 60%以上,说明这些性状在不同玉米品种中表现差异较大且易受环境影响,性状改良潜力较大,而生育期、出籽率的变异系数相对较小,这类性状相对稳定,较能反映玉米品种的稳产性,在玉米高产育种中,出籽率是可以直接进行选择的性状,这与张亚菲等<sup>[6]</sup>、钱双宏等<sup>[7]</sup>的研究结果基本一致。

对玉米品种进行科学有效评价,需要对各性状指标进行综合分析。主成分分析法通过降维思维,在最大程度保留原有数据信息的前提下,将原有多指标转化为个数较少的综合指标,可以避免重复信息干扰、排除主观因素影响。聚类分析是以各主成分的综合得分作为评价品种的新指标,采用平方欧式距离和平均联接(组间)构建系统聚类图,将不同品种进行归类,从而衡量各品种的优劣。近年来,主成分分析和聚类分析方法已被广泛应用在种质资源及作物杂交组合综合评价相关研究中<sup>[8-13]</sup>。

笔者选取 14 个农艺性状作为评价 33 个玉米品种的指标,通过主成分分析提取前 5 个主成分,累计贡献率达到 77.058%,将 14 个农艺性状评价指标转换成 5 个主成分指标,并对综合得分进行排序,然后应用欧式距离对 33 个玉米品种进行系统聚类分析,初步筛选出金园 23 和吉农玉 1 898 两个优异的玉米品种。较优品种聚在同一类群,间接验证了主成分分析和聚类分析结果的正确性和一致性,可为筛选适宜东北春玉米区种植的优良玉米品种及推广应用提供科学依据。

#### 参考文献

- [1]冯艳飞,杨威,任国信,等.黑龙江省部分玉米杂交种的综合评价[J].作物杂志,2021(4):46-50.
- [2]张召星,肖春华,邹楠,等.玉米品种高产稳产的综合评价研究[J].石河子大学学报(自然科学版),2017,35(4):425-430.
- [3]倪正斌,朱俊凯,孙红芹,等.22 个玉米杂交组合农艺性状的主成分和动态聚类分析[J].大麦与谷类科学,2017,34(5):18-22,37.
- [4]王江民,肖植文,张建华,等.聚类分析法在 28 个玉米杂交种筛选中的应用[J].西南农业学报,2014,27(2):480-484.
- [5]李洪,王瑞军,王彧超,等.玉米杂交组合主要农艺性状与产量的多重分析[J].中国农学通报,2018,34(22):31-36.
- [6]张亚菲,刘松涛,曹文梅,等.黄淮海夏玉米品种主要性状遗传多样性研究[J].种子,2021,40(4):96-100.

表 4 28 个参试玉米品种的主成分因子综合得分及排名

编号	主成分因子得分					综合得分	排序
	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$Y_5$		
1	-1.476	-0.953	1.962	0.814	-0.875	-0.291	20
2	-1.375	-0.876	-0.029	-0.295	0.196	-0.547	27
3	-0.332	0.121	1.157	1.754	1.842	0.423	11
4	1.284	1.821	2.338	1.533	1.480	1.272	1
5	-1.021	0.193	-1.434	0.722	-0.045	-0.353	23
6	1.557	0.838	0.781	0.167	-0.177	0.677	4
7	-2.06	1.485	1.178	-0.801	-2.310	-0.387	25
8	3.36	-1.874	1.389	-0.619	0.027	0.668	5
9	4.461	-2.215	1.265	-0.523	-0.645	0.837	3
10	-1.201	1.627	0.129	2.641	-0.374	0.251	13
11	1.516	0.363	-1.348	-0.226	-0.173	0.253	12
12	-3.096	-2.856	0.287	-0.621	0.023	-1.378	33
13	-1.642	1.181	2.262	-0.729	0.289	0.028	16
14	-1.459	1.542	-0.829	-1.281	-2.754	-0.571	29
15	1.025	-1.309	0.551	-0.567	-1.338	-0.064	17
16	-2.004	0.028	0.257	0.105	1.252	-0.382	24
17	0.861	0.984	0.877	-0.477	1.053	0.561	8
18	-0.165	1.167	-0.820	-2.484	0.860	-0.134	18
19	3.762	2.094	-1.299	0.136	-0.133	1.219	2
20	2.935	-2.271	2.090	0.335	-0.110	0.665	6
21	-1.906	-0.637	0.783	-0.193	0.411	-0.509	26
22	-1.956	-0.080	0.579	-1.914	1.255	-0.562	28
23	-1.947	-0.943	-1.839	0.207	-0.317	-0.94	31
24	0.849	0.630	-0.161	-0.558	-0.844	0.194	15
25	-0.404	-0.320	-2.048	1.492	-0.258	-0.300	21
26	2.846	1.371	-3.235	-1.231	1.268	0.550	9
27	0.488	-4.648	-2.363	-0.541	-0.021	-1.104	32
28	-0.530	0.435	-0.591	-0.603	-0.671	-0.257	19
29	0.870	1.690	-0.961	0.808	-0.627	0.452	10
30	-1.460	-0.728	-0.377	1.596	0.898	-0.330	22
31	0.575	2.060	-0.244	1.278	0.225	0.656	7
32	-0.858	1.805	1.656	-0.711	-0.346	0.222	14
33	-1.495	-1.727	-1.963	0.787	0.938	-0.819	30

[7]钱双宏,蔡世昆,王序英,等.18个玉米杂交组合农艺性状与产量的多元分析[J].天津农业科学,2019,25(11):22-28.

[8]贺礼英,尹成杰,黄守程,等.菜用大豆主要农艺性状的相关性、聚类及主成分分析[J].浙江农业学报,2018,30(1):50-57.

[9]何文,张秀芬,郭素云,等.基于主成分分析和聚类分析对22份马铃薯种质的综合评价[J].种子,2021,40(3):80-86.

[10]樊树雷,李和孟,邱宝财,等.基于主成分及聚类分析的10个杨梅品种生产性状综合评价[J].江苏农业科学,2020,48(14):



# 应用灰色系统关联分析法 对玉米新配组合的综合评价

谢文锦 李方明 李 宁 杨海龙 付 俊 张中伟 高旭东 丰 光

(丹东农业科学院 辽宁凤城 118109)

**摘要:**应用灰色系统关联分析法对新配玉米组合进行多农艺性状综合评价分析,结果表明,对产量和农艺性状灰色关联度分析可知,单穗粒重与产量关联度最大,轴粗与产量关联度最小。结合各组合性状构建最优品种,利用 12 个农艺性状权重计算各组合与最优品种的灰色关联度,筛选出 10 个综合性状好的、接近最优品种的新配组合。

**关键词:**玉米;农艺性状;产量;灰色关联度

玉米是世界上重要的粮食作物之一,其产量及品质对粮食安全和畜牧发展有着重要的意义与非凡的影响<sup>[1]</sup>。玉米的大多数农艺性状是由多个基因控制、互相影响的数量性状,因此有侧重性的选择玉米性状指标,对玉米育种和品种综合评价有重要的意义。灰色关联度分析法是一种基于灰色系统的多因素统计分析方法,在玉米育种中可以利用此方法来解决产量、品质与多种农艺性状的复杂的关联程度,是一种被证实的可靠性较强的分析方法<sup>[2]</sup>。

本研究利用灰色关联度分析法对院内 32 个新选配的玉米杂交组合主要农艺性状和产量进行遗传变异和相关性分析,以全面评价新选杂交组合的综合特性,为新组合的进一步利用提供重要的理论依据。

## 1 材料与方法

### 1.1 试验材料

以本院新配的 32 个玉米组合为试验材料(代号为 DP1~DP32),以郑单 958 作为对照品种(CK),于 2021 年春季在院内试验地播种。试验采用随机区组设计,设置 3 次重复,5 行区,行长 6.5 m,密度为

4 500 株/亩。主要调查性状为株高(X1)、穗位高、单穗粒重(X3)、穗粗(X4)、轴粗(X5)、穗长(X6)、秃尖长(X7)、穗行数(X8)、行粒数(X9)、出籽率(X10)、百粒重(X11),并计算穗位系数(X2)和产量(X12)。

### 1.2 数据分析

试验数据利用 R 语言和 DPS 7.5 进行统计分析。

灰色关联度分析依据邓聚龙<sup>[3]</sup>的方法,将新配组合产量作为母序列 X0,其余 11 个农艺性状为比较数列,计算产量与主要农艺性状的关联度。同时以试验组合为参考设定最优品种,计算各新组合与最优品种的关联度。

## 2 结果与分析

### 2.1 新组合不同农艺性状的相关分析

利用 R 语言的 PerformanceAnalytics 程序包对 12 个性状指标进行相关性作图分析,由附图可知,性状之间存在不同程度的相关性。株高和穗位系数显著负相关,且与穗长、秃尖呈显著正相关;穗位系数与株高、穗长和秃尖长呈显著负相关;单穗粒重与穗粗、穗长、行粒数、出籽率和产量显著相关;穗粗与轴

作者简介:谢文锦(1987-),女,硕士,助理研究员,从事玉米区域试验工作。E-mail:xiewenjia402@sina.cn

170-173.

[11]彭嘉熹,周仲华,马肖,等.陆地棉主推品种种质资源的评价与分析[J].湖南农业大学学报(自然科学版),2021,47(1):23-29.

[12]孙道旺,王艳青,洪波,等.云南冬播燕麦的农艺性状主成分

和聚类分析[J].作物杂志,2020(5):80-87.

[13]王桂梅,邢宝龙.芸豆品种主要农艺性状的主成分分析和聚类分析[J].种子,2021,40(2):76-79,85.