

药用印楝表型选择的因子分析及综合评价

彭兴民, 吴疆翀, 王有琼, 郑益兴, 马李一, 张燕平

(中国林业科学研究院资源昆虫研究所, 云南 昆明 650224)

摘要:以因子分析和聚类分析作为药用印楝表型选择的分析评价方法,对药用印楝表型选择的分析评价方法进行实证研究,结果表明:性状间呈显著或极显著相关;特征根 >1 的4个公因子累计方差贡献率为83.53%,第1、第2、第3和第4公因子方差贡献率分别为31.33%、22.98%、15.11%和14.10%,种仁质量(0.995 11)、种子质量(0.956 76)、种子长(0.780 12)、印楝素A+B总含量(0.866 67)、印楝素B含量(0.819 34)、印楝素A含量(0.680 84)、地径(0.939 04)、株产果实数(0.722 82)、印楝素B与印楝素A比(-0.833 60)是表型选择分析评价中的典型代表变量;Mantel检测发现,聚类分析能有效的表示33个基因型样株表型性状的相似(异)性($r = 0.81, p = 1.00$);各聚类群的性状特征概括为:种子印楝素A含量优异类、种子印楝素B含量优异类、种子高印楝素AB含量类、粒大饱满类、树体大结实多类和无优异特性类;优树的综合评价标准:综合得分 ≥ 0.59 的为“好”,综合得分为 $0.59 \sim -0.19$ 的为“中”,综合得分 ≤ -0.19 的为“差”,按此标准,筛选出“好”、“中”、“差”优树各4株、16株、12株。与传统方法相比,新方法可以更全面、科学、客观、便捷、直观地分析表型选择的综合选择指标,综合评价优树的优劣。

关键词:印楝;表型选择;因子分析;聚类分析;综合评价

中图分类号:S722.5 S759.3 文献标识码:A

Factor Analysis and Comprehensive Assessment of Phenotypic Selection of Official Neem Superior Trees

PENG Xing-min, WU Jiang-chong, WANG You-qiong, ZHENG Yi-xing, MA Li-yi, ZHANG Yan-ping*

(Research Institute of Resource Insects, Chinese Academy of Forestry, Kunming 650224, Yunnan, China)

Abstract: In order to find analysis and assessment methods of phenotypic selection of officinal neem superior trees, and to empirically evaluate the officinal neem resources and to finally provide evidences and guidance in officinal neem breeding, 12 characters correlated with quality and yield in 32 primarily selected superior trees and control were tested. The results are as follows: among characters, at least one was significantly or extremely significantly correlated with another. Accumulative contribution rate of 4 common factors (with latent root >1) was 83.53%, their contribution rate was 31.33% (the first PC, seed shape), 22.98% (the second PC, azadirachtin content), 15.11% (the third PC, seed yield), 14.10% (the fourth PC, ratio of azadirachtin ingredient), respectively. And typical example of variables in analysis and assessment of phenotypic selection were: kernel weight (0.995 11), seed weight (0.956 76), seed length (0.780 12), content of azadirachtin A + B (0.866 67), content of azadirachtin B (0.819 34), content of azadirachtin A (0.680 84), ground diameter (0.939 04), seed yield per tree (0.722 82), ratio of azadirachtin B to A (-0.833 60). Mantel test showed that cluster analysis could reveal the phenotypic differences of 33 trees ($r = 0.81, p = 1.00$). Cluster groups could be categorized as: type of high azadirachtin content, type of full size of seeds and kernels, type of larger tree and high seed yield, and type of normal. Criterion of comprehensive assessment of primarily selected superior trees: “Superior” (with score ≥ 0.59), “Me-

收稿日期: 2015-01-05

项目基金: 国家“十一·五”科技支撑计划项目(2006BAD18B0302); 国家“十二·五”科技支撑计划项目(2012BAD21B04) 和国家自然科学基金(31270710)

作者简介: 彭兴民(1962—),男,高级工程师. 主要研究方向:植物引种、种质创新与新品种选育. E-mail: pengong007@21cn.com

dium" (with score $-0.19 \sim 0.59$), "Poor" (with score ≤ -0.19). According to this criterion, primarily selected superior trees were divided into 4 "Superior", 16 "Medium", and 12 "Poor" trees. With factor analysis and cluster analysis, analysis and assessment of phenotypic selection of officinal neem superior trees can be directly and easily performed and judged, the results could provide evidences and guidance in officinal neem breeding.

Key words: neem; officinal superior trees; phenotypic selection; factor analysis; comprehensive assessment

在实现树木良种化的过程中,有2次选择,即从自然树木到优良类型(表型选择)、再到原种(表型测定中还有1次再选择),这是从遗传性上解决质的过程,表型选择的分析及评价是实现树木良种化不可缺少的重要环节,在树木育种中具有重要的意义。印楝(*Azadirachta indica* A. Juss.)是5种直接用于加工农药的植物之一^[1],对于全球防治病虫害、保障食品和环境安全具有重要意义^[2]。我国印楝规模化引种已达十余年,药用印楝遗传改良工作正逐步开展^[3]。研究认为,印楝栽培种群是一个变异的群体,遗传改良应注重优树变异的选择和利用^[4-7]。因此,研究表型选择的分析评价方法,具有重要的理论与应用价值。

本课题在对云南12个乡镇31个选优样区的印楝资源进行调查的基础上,以选优林分个体适应性(病虫害、旱害和冻害)、树木结构(树高、枝下高、地径和冠幅)、生物产量(单株结实量、种子长、种子宽、种子粒质量、种仁粒质量和出仁率)和经济产量(印楝素组分及其含量)等为评价指标,采用传统的表型选择分析评价方法筛选了一批优树^[8];但传统的表型选择分析评价方法难以完整体现药用印楝表型选择所包含的属性,需要重新建立一套适合药用印楝表型选择分析评价的方法。国外,印楝表型性状在种源、群体水平上的变异已有研究^[9-10],但药用印楝表型选择以及表型选择分析评价方法的建立等研究均未见报道。因子分析和聚类分析都是多对象的多元变量分析,目的是进行变量的简化和排序或分类^[11]。在植物领域,因子分析和聚类分析已广泛应用于植株或品种间数量性状,尤其是品质性状的分析评价^[12-17]。本研究在前期工作的基础上,以32株药用印楝初选优树及对照的12个品质和产量相关性状为研究对象,采用因子分析和聚类分析构建新的药用印楝表型选择分析评价体系,对所建立的药用印楝表型选择分析评价方法进行实证研究,旨在重新建立一套适合药用印楝表型选择的分析评价方法,并在理论和实践上为分析评价药用印楝资源和指导药用印楝育种提供新的思路和方法。

1 材料与方法

1.1 试验材料

在印楝云南分布区进行药用优树选择,采用对比木法选出233株药用候选树,其中,印楝素总含量(azAB) $> 0.80\%$ 的候选树32株,为初选优树^[8]。选优样区位于 $23^{\circ}06' \sim 25^{\circ}51'N$, $101^{\circ}52' \sim 103^{\circ}11'E$,海拔197~1325 m,年平均气温 $21.8 \sim 24.5^{\circ}C$,年平均降水量634~805.1 mm。选优林分造林地均为宜林荒山,土壤为燥红土,零星种植,株行距 $4 m \times 4 m$,面积一般不少于 $10 hm^2$,树龄5~7 a。初选优树所在林分的平均木的总体平均值为对照(CK)。初选优树的名称见表1。

1.2 测定项目与方法

12个品质和产量相关性状是在前文^[6]基础上选取的。用游标卡尺量测地径(Dz)。株产果实数(Gss)的估测方法:在东、西两面树冠的上、中、下3个位置各选3个能代表结果状况的样枝,统计果实数,计算平均每个样枝的果实数,统计植株的结果枝数。

株产果实数 = 结果枝数 \times 样枝果实数

果实成熟中期,8:00—12:00在树冠南面的中部采样,分株采收、存放,约500粒 \cdot 株⁻¹,过熟果(果皮变黄)不采,只采青色果,采下青色果存放2 d,拣变黄变软的果实,除去青色、硬果,人工清洗,自然风干,保存于冰箱冷藏室($4^{\circ}C$ 左右)待用。采用GB/T3543.7—1995的百粒四分法随机抽取30粒种子,用游标卡尺测量种子长(Sl)、种子宽(Sw),测量精度0.01 cm,计算种子宽长比(SwL);用1/10000天平称量种子粒质量(St)和种仁粒质量(Kt),称量精度为0.01 g,计算出仁率(KSt)。用宗乾收等^[18]提出的“印楝种仁中印楝素含量的快速液相色谱分析”法分析种子印楝素A含量(azA)和印楝素B含量(azB),计算印楝素A+印楝素B总含量(azAB)和印楝素B与印楝素A比(azBA)。性状检测人员和检测仪器固定。

1.3 数据处理

首先采用Microsoft Office Excel2003软件对原始

数据进行描述统计和隶属函数法(公式 1)转换^[16];转换后的数据采用 SAS 9.0 system for windows(SAS Institute Inc., Cary, NC, USA.) 软件进行相关分析和因子分析^[19]。因子分析提取公因子的方法为迭代主因子法,正交方差最大旋转,得到各公因子的特征值、贡献率 E_j 、累计贡献率以及因子载荷阵、公因子分值 F_{jn} ;综合评价指标得分 D_n 的计算以相应公因子的贡献率 E_j 为权重,通过对各因子得分值 F_{jn} 进行加权得到(公式 2)。

$$U_{in} = \frac{X_{in} - X_{imin}}{X_{imax} - X_{imin}} \quad (1)$$

$$D_n = \sum_{j=1}^m F_{jn} \times E_j \quad (2)$$

式中: U_{in} 指第 n 个性状第 i 个指标的原始数据经转换后的隶属函数值; X_{in} 指第 n 个性状第 i 个指标的原始测定结果; X_{imax} 和 X_{imin} 分别指性状组中第 i

个指标的最大和最小值; D_n 为因子分析法得到的各性状的综合分值; F_{jn} 为第 n 个性状第 j 个特征根 > 1 的公因子的分值; m 为特征根 > 1 的公因子的个数; E_j 为第 j 个公因子的方差贡献率。

以提取公因子值为指标,采用 NTSYS PC 2.11e 软件(Applied Biostatistics Inc., Setauket, USA) 计算 DIST 距离系数,并用 UPGMA 聚类算法进行系统聚类^[20],最后用 Mantel 检测分析 33 个基因型样株提取公因子综合得分的空间距离和距离矩阵间的同表象相关性,以验证聚类是否能有效表示 33 个基因型样株表型性状的相似(异)性。

2 结果与分析

2.1 优树表型性状测定与结果分析

32 株印楝药用初选优树及对照(CK)表型性状的测定结果(表 1)显示:与对照相比,初选优树在印

表 1 初选优树表型性状测定结果

优树		性状											
编号	名称	azA/%	azB/%	azAB/%	azBA	Sl/cm	Sw/cm	Swl	St/g	Kt/g	KSt/g	Gss/粒	Dz/cm
1	Ww0401	0.78	0.20	0.97	0.25	1.39	0.59	0.43	0.16	0.08	0.482 3	2 650	11.80
2	Ww0402	0.71	0.20	0.91	0.28	1.15	0.60	0.52	0.13	0.07	0.490 9	2 760	12.30
3	Ww0410	0.65	0.16	0.82	0.25	1.32	0.61	0.47	0.17	0.08	0.475	3 980	15.20
4	Ww0411	0.68	0.17	0.85	0.25	1.29	0.58	0.45	0.14	0.07	0.457 9	3 690	14.90
5	Ww0418	0.70	0.19	0.89	0.26	1.34	0.63	0.47	0.16	0.07	0.459 2	3 400	11.60
6	Ww0421	0.64	0.29	0.93	0.46	1.26	0.65	0.52	0.14	0.06	0.445 4	2 170	11.00
7	Wx0416	0.65	0.19	0.85	0.30	1.60	0.77	0.48	0.25	0.13	0.508 1	5 500	19.60
8	Wx0423	0.72	0.18	0.90	0.25	1.69	0.66	0.39	0.24	0.12	0.508 3	5 200	18.30
9	Nms0503	0.63	0.18	0.81	0.29	1.48	0.70	0.47	0.22	0.12	0.517 9	2 900	13.60
10	Nms0505	0.69	0.17	0.86	0.25	1.31	0.71	0.54	0.19	0.10	0.513	2 980	13.70
11	Nms0506	0.69	0.18	0.87	0.26	1.22	0.66	0.54	0.19	0.09	0.492	3 280	14.20
12	Nms0509	0.71	0.20	0.90	0.28	1.41	0.64	0.45	0.21	0.11	0.526 6	3 540	13.80
13	Zz0501	0.65	0.17	0.82	0.26	1.29	0.64	0.50	0.13	0.06	0.481 5	8 520	18.50
14	Ld0501	0.63	0.21	0.84	0.34	1.18	0.64	0.54	0.13	0.06	0.446 3	7 520	23.00
15	Ld0502	0.61	0.18	0.79	0.30	1.38	0.71	0.51	0.23	0.12	0.522 1	6 780	24.00
16	Ld0505	0.85	0.24	1.09	0.28	1.48	0.65	0.44	0.17	0.08	0.455 8	6 050	20.50
17	Ld0507	0.70	0.15	0.84	0.21	1.24	0.63	0.50	0.14	0.07	0.460 1	5 100	14.50
18	Ld0509	0.72	0.18	0.90	0.25	1.38	0.71	0.51	0.18	0.08	0.456	4 600	16.10
19	Ld0510	0.65	0.20	0.85	0.31	1.66	0.67	0.40	0.29	0.13	0.454 2	5 020	16.60
20	Ld0511	0.70	0.22	0.92	0.32	1.48	0.68	0.46	0.14	0.07	0.462 4	5 280	17.80
21	Ld0513	0.69	0.17	0.87	0.25	1.27	0.66	0.52	0.15	0.07	0.464 5	5 110	18.40
22	At0515	0.68	0.27	0.94	0.39	1.29	0.57	0.45	0.12	0.05	0.438 4	4 500	17.90
23	At0516	0.73	0.17	0.90	0.23	1.30	0.58	0.45	0.13	0.06	0.466 6	4 310	14.50
24	At0517	0.68	0.17	0.85	0.25	1.48	0.60	0.41	0.18	0.10	0.539 8	3 780	14.60
25	At0518	0.64	0.19	0.83	0.30	1.44	0.66	0.46	0.21	0.10	0.497 6	3 200	16.40
26	Dhg0504	0.66	0.17	0.83	0.26	1.19	0.57	0.48	0.13	0.06	0.477 4	6 500	16.10
27	Dhg0506	0.67	0.16	0.83	0.24	1.20	0.63	0.61	0.18	0.09	0.478 9	7 500	17.20
28	Dhg0507	0.73	0.22	0.95	0.30	1.33	0.60	0.46	0.16	0.08	0.480 5	6 400	16.70
29	Dhg0510	0.70	0.20	0.90	0.29	1.17	0.61	0.53	0.13	0.06	0.470 5	8 200	24.10
30	Ld0612	0.64	0.17	0.81	0.27	1.14	0.59	0.52	0.09	0.03	0.347 1	5 850	19.00
31	Ld0642	0.64	0.15	0.79	0.24	1.15	0.65	0.57	0.14	0.08	0.520 7	5 900	23.00
32	Ww0603	0.63	0.17	0.80	0.26	1.21	0.67	0.56	0.13	0.07	0.483 4	2 110	13.50
33	CK	0.47	0.14	0.61	0.29	1.39	0.67	0.48	0.22	0.10	0.454 5	2 335	14.4

注:azA,印楝素 A 含量(%); azB,印楝素 B 含量(%); azAB,印楝素 A + B 总含量(%); azBA,印楝素 B 比印楝素 A; Sl,种子长(cm); Sw,种子宽(cm); Swl,种子宽长比; St,种子粒质量(g); Kt,种仁粒质量(g); KSt,出仁率(%); Gss,株产果实数(粒); Dz,地径(cm)。

楝素组分含量和树体结实量方面表现较好,而在种子(仁)形态(大小、质量)方面则表现一般,说明32株印楝药用初选优树在各项表型指标中表现出不同程度的优良特性,对上述初选优树进行选优,必须依据样品在12项表型指标中的综合表现。根据药用印楝的特性和选优目标,在表型选择时各表型指标的衡量标准不同:印楝素A、B含量越高越好,含量越高,表明品质越好;种子(仁)越大越好、质量越重越好,种子(仁)大、重,表明种子(仁)粒大饱满,产量高;树体和地径越大越好,树体和地径大小与产量呈正相关,树体和地径大表明产量高。另外,不同指标的计量单位或量纲不同,数据数量级也不一致,不

便于进行数据的分析,因此,在进行数据分析时需对原始数据进行转化。

2.2 优树性状变异的相关性

经隶属函数法转换的数据进行简单相关性分析,相关系数矩阵见表2。从表2可以看出:12个性状都至少与一个其它性状呈显著或极显著相关,其中,印楝素A和印楝素B与印楝素AB总含量、种子长和种子宽与种子质量和种仁质量、种子质量与种仁质量、地径与株产果实数,均达到极显著水平,呈极显著正相关,结果表明,各表型性状彼此相关,并不独立,表明印楝药用优树表型是一个复杂的综合评价指标,有必要对其使用因子分析。

表2 初选优树性状的简单相关系数矩阵

	azA	azB	azAB	azBA	Sl	Sw	Swl	St	Kt	KSt	Gss	Dz
azA	1.000 00	NS	* *	NS	NS	NS	NS	NS	NS	NS	NS	NS
azB	0.338 29	1.000 00	* *	* *	NS	NS	NS	NS	NS	NS	NS	NS
azAB	0.920 76	0.674 44	1.000 00	NS	NS	NS	NS	NS	NS	NS	NS	NS
azBA	-0.214 52	0.841 06	0.179 50	1.000 00	NS	NS	NS	NS	NS	NS	NS	NS
Sl	0.111 64	0.116 58	0.143 05	0.066 21	1.000 00	* *	* *	* *	* *	NS	NS	NS
Sw	-0.247 37	-0.090 76	-0.212 13	0.079 18	0.462 96	1.000 00	NS	* *	* *	NS	NS	NS
Swl	-0.276 82	-0.214 13	-0.298 97	-0.050 44	-0.747 28	0.170 03	1.000 00	*	NS	NS	NS	NS
St	-0.181 44	-0.122 99	-0.185 49	-0.007 62	0.791 55	0.629 89	-0.370 80	1.000 00	* *	* *	NS	NS
Kt	-0.147 06	-0.186 58	-0.186 15	-0.090 84	0.747 86	0.645 08	-0.311 32	0.960 02	1.000 00	* *	NS	NS
KSt	0.035 35	-0.211 63	-0.059 05	-0.229 52	0.319 35	0.340 35	-0.087 36	0.486 00	0.673 77	1.000 00	NS	NS
Gss	0.104 32	-0.053 88	0.068 09	-0.087 37	-0.147 52	-0.068 73	0.203 98	-0.160 53	-0.176 86	-0.131 45	1.000 00	* *
Dz	-0.055 72	-0.023 66	-0.040 44	0.040 89	-0.045 81	0.164 36	0.200 11	0.015 26	-0.011 18	-0.044 79	0.804 75	1.000 00

注: $r_{0.05(32)} = 0.349$; $r_{0.01(32)} = 0.449$ 。*、*、*分别表示在5%和1%水平显著性;NS表示差异不显著。

2.3 影响优树间性状差异的因子数及其解释

迭代主因子分析12项指标的前4个公因子(特征根>1)、方差贡献率 E_j 、累计方差贡献率和因子载荷矩阵见表3。从表3可以看出:前4个综合指标或公因子的累计贡献率达83.53%,即这4个公因子所含信息占总体信息的83.53%,表明前4个综合指标能代表12个单项指标的绝大部分信息,且具有较明显的生物学意义,可以分别用这4个公因子对33样株表型选择指标进行概括分析评价。第1公因子(PC1)贡献率为31.33%,决定第1公因子大小的主要是种仁质量(Kt)、种子质量(St)和种子长(Sl),其特征向量所凝聚的生物学信息主要是种子形态因素或产量组分;其向量间的关系表明,种子粒大饱满,产量高。第2公因子(PC2)贡献率为22.98%,决定第2公因子大小的主要是印楝素总含量(azAB)、印楝素B含量(azB)和印楝素A含量(azA),其特征向量所揭示的生物学信息主要是印楝素组分及其含量

构成因子;其向量间关系表明,种子印楝素A含量(azA)、印楝素B含量(azB)和印楝素A+印楝素B总含量(azAB)高,种子品质好。第3公因子(PC3)贡献率15.11%,决定第3公因子大小的是地径(Dz)和株产果实数(Gss),其特征向量所凝聚的生物学信息主要是树体结实量因子;其向量间关系表明,株产果越多,产量越高,而地径(Dz)大小与产量高低正相关,地径(Dz)越大,产量越高。第4公因子贡献率14.10%,决定第4公因子(PC4)大小的主要是印楝素B含量与印楝素A含量比(azBA),其特征向量所凝聚的生物学信息主要是印楝素B含量与印楝素A含量比率关系,其大小反映印楝素组分的变化,为印楝素组分比因子。第1和第3公因子都与产量相关,可以大致概括为产量因子;第2、第4公因子都与品质相关,可以大致概括为种子品质性状因子。

从以上结果可以看出:因子分析可获得33样株

的相关矩阵的特征值、贡献率 E_j 、累计贡献率、因子载荷阵,能探求并解释影响药用印楝表型选择分析评价的关键因子,验证表明,因子分析用于药用印楝

表型选择分析可行。因此,药用印楝表型选择适合使用因子分析进行分析。

表 3 影响初选优树间性状差异的因子数及其解释

性状和指标	因子				变量共同度
	第 1 公因子	第 2 公因子	第 3 公因子	第 4 公因子	
azA	-0.232 89	0.680 84	-0.044 32	0.646 17	0.914 309 43
azB	-0.255 82	0.819 34	0.214 97	-0.435 19	0.965 696 75
azAB	-0.280 28	0.866 67	0.067 40	0.333 17	0.922 042 36
azBA	-0.118 37	0.458 62	0.293 41	-0.833 60	0.967 618 07
Sl	0.780 12	0.446 50	0.082 42	0.065 97	0.866 632 65
Sw	0.599 79	-0.089 07	0.245 24	-0.137 70	0.645 782 85
Swl	-0.332 47	-0.500 65	0.138 35	-0.112 24	0.600 872 89
St	0.956 76	0.101 26	0.119 86	-0.033 92	0.915 706 56
Kt	0.995 11	0.057 70	0.096 89	0.052 57	0.953 321 30
KSt	0.539 78	-0.013 87	-0.025 81	0.187 11	0.489 235 72
Gss	-0.250 70	-0.157 13	0.722 82	0.308 01	0.883 523 09
Dz	-0.088 45	-0.201 41	0.939 04	0.180 34	0.898 653 02
特征值	3.759 417 99	2.758 037 24	1.813 461 41	1.692 478 12	
贡献率/%	31.33	22.98	15.11	14.10	
累计贡献率/%	31.33	54.31	69.42	83.53	

注:性状同表 1。

2.4 各优树的因子得分及排名

因子分析还能获得各公因子分值 F_{jn} 。根据各公因子分值 F_{jn} 可对 33 样株在每个公因子上进行排序。33 样株在每个公因子上的得分和排序情况(表 4)表明:33 样株在 4 个公因子上的得分和排序不一致。

以上结果显示:因子分析可对 33 样株在每个公因子上进行排序。因此,表型选择适合使用因子分析进行评价;但各优树在 4 个公因子上的得分和排序不一致,表明用单个公因子评价药用印楝优树会得出不同甚至相反的结论。因此,必须进行综合评价。

2.5 各优树的分布

为探明各优树在 4 个公因子中的分布则需作聚类分析。以 33 样株的 4 个公因子值为指标,进行系统聚类分析,聚类结果见图 1。图 1 反映了优树间的相似性关系,欧式平均距离为 2.224 ~ 0.176。按欧式平均距离 1.712 的阈值把 33 个基因型聚为 4 类:即由 Ww0421、Ld0505 和 CK 3 株自成体系的个类(I 类、II 类和 IV 类)和由其余 30 株优树聚成的 III 类。按欧式平均距离 1.200 的阈值把 III 类分为 4 个亚类:Wx0416、Wx0423、Ld0502、Ld0510 组成 III-1 亚类, Ld0511、At0515、Dhg0507 组成 III-2 亚类, Zz0501、Dhg0506、Ld0642、Ld0501、Dhg0510、Ld0612 组成 III-4 亚类,其余 17 株优树组成 III-3 亚类。除分

“类”和“亚类”外,还可按欧式平均距离 0.688 的阈值把各亚类分为次亚类,从次亚类中找出具有优异特性的单株。

Mantel 检测发现:表象相关系数 $r = 0.81$, $p = 1.00$, 表明 33 样株综合得分的 DIST 距离系数的空间距离和距离矩阵间全相关,聚类分析能有效的表示 33 样株表型性状的相似(异)性。各聚类群的性状特征概括为: I 类(Ww0421)和 II 类(Ld0505)分别为种子印楝素 B 含量、印楝素 A 含量优异类, III-1 亚类为粒大饱满类, III-2 亚类为高印楝素总含量(azAB)类, III-3 亚类为树大结实多类,而 III-4 亚类与 IV 类(对照)同类,为无优异特性类(表 5)。

聚类分析可以直观地揭示各印楝药用优树间表型性状差异状况,能更具直观、简便地区分出各优树自然类型分类的特点;所以,药用印楝表型选择的分析与评价适合使用聚类分析。

2.6 优树的综合评价

以 33 样株因子分析相应公因子的贡献率 E_j 为权重,通过对各因子得分值 F_{jn} 进行加权得到综合评价指标得分 D_n (公式 2),并按综合得分进行排序(表 4)。根据 33 样株的分布(图 1)及其综合得分(表 4),可计算出各聚类样株的平均综合得分(表 5)。各类群平均综合得分值大小或综合表现排序: II 类 > III-1 类 > III-2 类 > I 类 > III-3 类 > III-4 类 >

表4 各初选优树因子得分及排名

优树		因子								综合评价	
		第1公因子		第2公因子		第4公因子		第4公因子			
编号	名称	得分	排名	得分	排名	得分	排名	得分	排名	得分	排名
1	Ww0401	-0.198 56	17	1.744 46	2	-0.493 40	22	-1.369 43	32	0.071 03	15
2	Ww0402	-0.961 20	30	0.294 64	11	-0.144 46	16	-1.121 76	29	-0.413 43	29
3	Ww0410	-0.184 11	15	-0.250 96	19	-0.727 43	28	-0.598 28	22	-0.309 62	25
4	Ww0411	-0.694 42	25	0.271 18	12	-0.650 73	26	-0.809 47	25	-0.367 71	27
5	Ww0418	-0.383 15	19	0.408 23	10	-0.195 85	17	-1.128 69	30	-0.214 97	22
6	Ww0421	-0.756 85	26	-0.535 19	25	3.592 69	1	-1.400 04	33	-0.014 66	16
7	Wx0416	2.205 26	1	-0.453 58	24	0.564 50	7	0.949 79	8	0.80589	3
8	Wx0423	1.857 86	3	1.210 08	3	-0.479 97	21	0.257 76	13	0.823 96	2
9	Nms0503	1.375 88	5	-0.671 43	26	0.064 07	13	-0.889 96	27	0.160 97	9
10	Nms0505	0.528 56	10	-0.387 45	23	-0.630 69	25	-0.561 78	21	-0.097 95	18
11	Nms0506	-0.026 21	14	-0.279 38	21	-0.421 96	20	-0.529 79	20	-0.210 87	21
12	Nms0509	0.895 39	6	0.663 91	6	-0.132 11	15	-0.746 68	24	0.307 85	8
13	Zz0501	-0.507 75	22	-0.278 53	20	-0.517 88	24	1.264 14	5	-0.123 09	19
14	Ld0501	-0.834 02	28	-0.825 17	29	1.157 02	3	1.740 35	3	-0.030 71	17
15	Ld0502	1.450 12	4	-1.092 03	31	0.227 04	10	1.775 26	2	0.487 99	5
16	Ld0505	0.277 29	11	2.738 18	1	0.616 52	6	1.154 93	6	0.972 11	1
17	Ld0507	-0.687 60	24	0.066 19	16	-1.360 72	33	-0.208 11	18	-0.435 16	30
18	Ld0509	0.219 76	12	0.218 00	14	-0.230 89	18	0.190 33	14	0.110 90	12
19	Ld0510	1.996 16	2	0.134 56	15	0.845 78	5	-0.074 51	16	0.773 61	4
20	Ld0511	0.125 75	13	0.546 02	9	1.061 73	4	0.396 03	12	0.381 14	6
21	Ld0513	-0.392 23	21	-0.087 04	18	-0.494 46	23	0.499 49	9	-0.147 17	20
22	At0515	-1.052 35	32	0.615 14	8	2.366 64	2	-0.119 36	17	0.152 43	11
23	At0516	-0.833 24	27	0.975 02	5	-1.008 60	31	-0.617 75	23	-0.276 50	24
24	At0517	0.724 70	8	0.640 16	7	-0.943 70	29	-0.856 54	26	0.110 79	13
25	At0518	0.844 12	7	-0.381 98	22	0.333 21	9	-0.520 78	19	0.153 60	10
26	Dhg0504	-1.004 10	31	-0.023 75	17	-0.703 25	27	0.152 71	15	-0.404 77	28
27	Dhg0506	-0.390 12	20	-0.690 89	27	-0.965 57	30	1.113 90	7	-0.269 83	23
28	Dhg0507	-0.193 10	16	1.134 75	4	0.413 78	8	0.411 74	11	0.320 84	7
29	Dhg0510	-0.866 84	29	0.258 46	13	0.154 86	12	2.097 77	1	0.107 00	14
30	Ld0612	-2.184 07	33	-0.768 09	28	0.161 17	11	0.476 86	10	-0.769 19	33
31	Ld0642	-0.343 83	18	-1.018 28	30	-1.115 06	32	1.385 49	4	-0.314 85	26
32	Ww0603	-0.646 13	23	-1.191 54	32	-0.383 88	19	-1.010 81	28	-0.676 78	31
33	CK	0.639 02	9	-2.983 71	33	0.041 58	14	-1.302 82	31	-0.662 87	32

CK类。由各类群样株综合得分的变幅看出:类群间样株并没有完全按综合得分大小分布,表明各类群的平均综合得分并不能作为优树实际表现的分类标

准,若按优树的实际表现进行分类,还需进行再选择。

表5 各聚类群性状的平均值及综合得分

性状	I	II	III				IV(CK)
			III-1	III-2	III-3	III-4	
azA/%	0.64	0.85	0.66	0.70	0.66	0.69	0.47
azB/%	0.29	0.24	0.19	0.24	0.18	0.18	0.14
azAB/%	0.93	1.09	0.85	0.94	0.83	0.86	0.61
Sl/cm	1.26	1.48	1.58	1.37	1.19	1.32	1.39
Sw/cm	0.65	0.65	0.70	0.62	0.63	0.64	0.67
St/g	0.14	0.17	0.25	0.14	0.13	0.17	0.22
Kt/g	0.06	0.08	0.13	0.07	0.06	0.08	0.10
Gss/粒	2 170.00	6 050.00	5 625.00	5 393.00	7 248.00	3 758.24	2 335.00
Dz/cm	11.00	20.50	19.63	17.47	20.80	14.42	14.40
综合得分	-0.01	0.97	0.82~0.49	0.38~0.15	0.31~-0.68	0.11~-0.77	-0.66
平均综合得分	-0.01	0.97	0.72	0.28	-0.16	-0.23	-0.66

注:性状同表1。

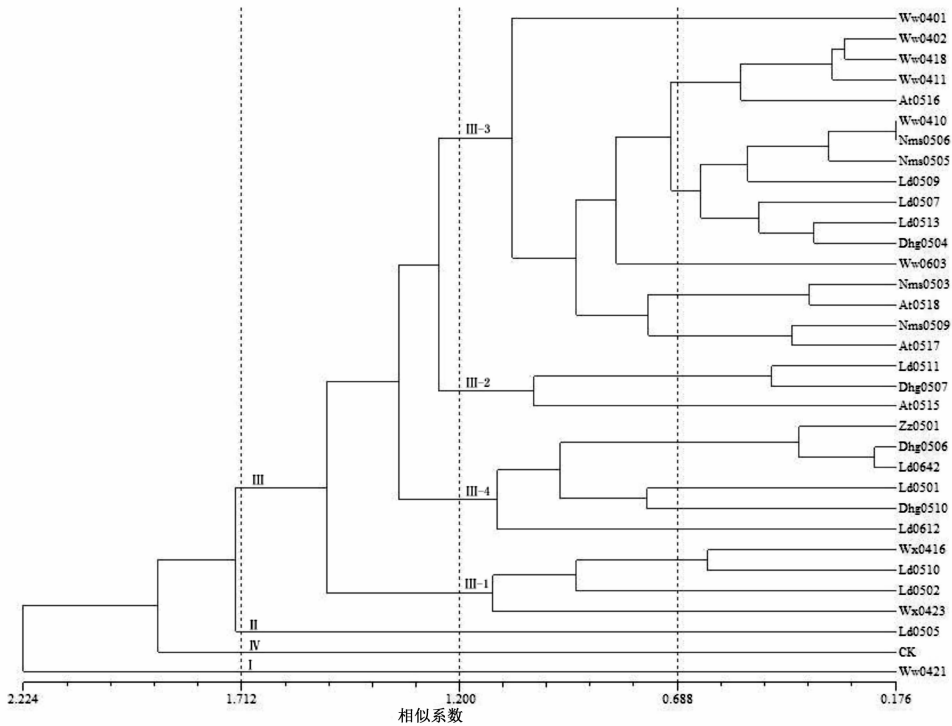


图1 32株药用印楝初选优树及对照的UPGMA聚类图(Mantel检测, $r=0.81, p=1.00$)

II类群(Ld0505)、III-1类群(Wx0416、Wx0423、Ld0502、Ld0510)和III-2类群(Ld0511、At0515、Dhg0507)8株综合排名均名列前茅,性状表现好,其余25株综合排名居后,性状表现较差。分别计算性状表现好的8株和性状表现较差的25株的平均综合得分,以此制定32株优树表型选择的综合评价标准:综合得分 ≥ 0.59 的为“好”,综合得分为 $0.59 \sim -0.19$ 的为“中”,综合得分 ≤ -0.19 的为“差”。按此标准,可筛选出表型选择“好”、“中”、“差”的各

类优树(表6)。由各类群的综合得分变幅可以看出:类群间优树综合得分按大小分布,表明按此标准进行分类能完全反映各优树的实际表现。

综合评价结果表明:因子分析结合聚类分析可计算出各聚类群优树的平均综合得分,以此制定药用印楝表型选择的综合评价标准,对优树作出综合评价。因此,表型选择适合使用因子分析和聚类分析进行综合评价。

表6 32株印楝药用初选优树综合评价

综合评价	综合得分	平均综合得分	样本量	样本分布情况
好	≥ 0.59	0.60 0.97 ~ 0.77	4	Ld0505、Wx0416、Wx0423、Ld0510
中	$0.59 \sim -0.19$	0.12 0.49 ~ -0.15	16	Ld0502、Ld0511、At0515、Dhg0507、Ww0421、Ww0401、Nms0503、Nms0505、Nms0509、Ld0509、At0517、At0518、Ld0501、Dhg0510、Ld0513、Zz0501
差	≤ -0.19	-0.41 -0.211 ~ -0.77	13	Ww0402、Ww0410、Ww0411、Ww0418、Nms0506、Ld0507、At0516、Dhg0504、Ww0603、Dhg0506、Ld0642、Ld0612、CK

注:平均综合得分值列各行下排为综合得分值。

3 小结与讨论

3.1 表型选择分析评价方法的构建

印楝株果实数、地径、树高、印楝素A、B总含量、印楝素A含量及印楝素B含量等6个表型性状是造成表型差异的主要表型性状^[6];因此,本研究主

要从种子品质与产量的角度对表型选择进行分析评价。本研究12个性状都至少与一个其它性状呈显著或极显著相关,表明药用印楝优树表型是一个复杂的综合评价指标,表型选择分析评价实质上是一个多目标的多元统计分析问题。定性描述、描述统计、单变量分析和专家打分等是过去表型选择分析

评价普遍采用的方法。彼此相关的指标采用上述传统分析评价方法进行分析评价,会导致各单项指标提供的信息发生重叠,不易得出简明的规律,不仅增加了分析问题的难度,而且会歪曲研究对象间的真实差异^[15,20]。前期研究^[8]也表明,采用 t 检验方法进行表型选择的分析评价,只能对优树与对比木间差异进行比较,获得对比木优良的单株,但入选单株(或优树)间的差异不能深入探究,难以完整体现药用印楝表型选择所包含的属性。综上,传统分析评价方法的视角单一,技术手段的科学性、客观性受到质疑,也较繁琐,所获结果是初步的,且不够直观。因此,上述传统分析评价方法不适于做表型选择的深入分析及评价。

多目标的多元统计分析进行变量的简化和排序或分类,基本方法有二,方法之一是推导变量的线性组合以概括原始数据集中的差异(R方式分析法),也就是基于从变量或对象间相关矩阵中抽取的特征向量和特征值进行变量的简化和排序;另一种方法是根据对象间不相似测度的方法(Q方式分析法),也就是基于对象间相似性或不相似性直接对对象进行简化和分类。基于从变量或对象间相关矩阵中抽取的特征向量和特征值的方法很多,目前主要采用主成分分析(PCA)。主成分分析采用的分析方法相对来说比较简单和粗糙,而因子分析采用的分析方法更加精细和深入^[21]。因子分析是基于从变量或对象间相关矩阵中抽取的特征向量和特征值进行变量的简化和排序,目的或结果是排序;聚类分析是根据对象间不相似测度进行变量的简化和分类,目的或结果是分类。结果展示上,因子分析有“序”无“类”,而聚类分析则是有“类”无“序”,两种分析结合能进行变量的简化和排序或分类。基于上述认识,本研究采用因子分析和聚类分析构建新的药用印楝表型选择分析评价体系。

3.2 新建表型选择分析评价方法的实证研究

原始数据直接用于分析或不同转换方法对分析结果有较大影响^[14,16]。本研究不同指标的计量单位(量纲)不同,数据数量级也不一致。因此,在使用因子分析进行分析评价时必须对原始数据作规范化处理,即区分评价指标的类型和评价指标的无量纲化处理。评价指标包括4种基本类型,不同指标类型数据有各自的转换公式^[22]。本研究指标类型为正相关型指标,故采用正相关型指标转换公式(隶属函数法)进行转换。

考察原始变量之间的相关性是因子分析的首要步骤,如果各变量之间是独立的,那么可能不适合使用因子分析^[19]。为此,对经隶属函数法转化的数据进行相关分析就成为因子分析的首要步骤。本研究的各表型性状彼此相关,故适合使用因子分析。如何构造因子变量和对因子变量进行命名解释,是因子分析的两个核心问题^[19]。提取特征根 >1 的公因子,获得相关矩阵的累计贡献率、因子载荷阵等值。根据累积方差贡献率阈值确认提取公因子的数目,即累积方差贡献率85%阈值来决定,方差贡献率越大,因子分析越有意义;因子载荷是第 i 个变量与第 j 个因子之间的相关系数,反映了第 i 个变量在第 j 个因子上的重要性,根据因子载荷大小确认因子的典型代表变量,因子的典型代表变量突出,即每个原始变量仅在一个公因子上有较大的载荷,而在其余的公因子上的载荷比较小。本研究提取到种子形态、印楝素组分及其含量、印楝素组分比和树脂结实量4个公因子,所含信息占总体信息的83.53%,表明前4个综合指标能代表12个单项指标的绝大部分信息,可以分别用这4个公因子对33样株表型选择进行概括分析评价。

本研究表明,因子分析还能获得33样株各公因子分值 F_{jn} 和公因子的贡献率 E_j ,根据各公因子分值 F_{jn} 可对33样株在每个公因子上进行排序;以相应公因子的贡献率 E_j 为权重,通过对各因子得分值 F_{jn} 进行加权得到综合评价指标得分 D_n (公式2),并按综合得分进行排序;以4个公因子值为指标,进行UPGMA系统聚类分析,获聚类图,可探明各样株在4个公因子中的分布。由各类群样株综合得分,分别计算性状表现好的8株和性状表现较差的25株的平均综合得分,以此制定32株优树表型选择的综合评价标准。按此标准,可筛选出表型选择“好”、“中”、“差”的各类优树。

聚类中引入公因子的目的是用尽可能少的公因子说明生物学的大部分信息,减少统计的复杂性^[23]。聚类分析能直观地揭示研究对象在各公因子中的分布情况,但由于各公因子在因子分析时的方差贡献率不同,所以在实际应用时,还要结合因子贡献率,协调好各公因子之间的侧重关系。以往研究,无聚类矩阵间的相关性检验;但按聚类分析程序要求,尚需再作Mantel检测,以验证聚类的代表性;所以,无Mantel检测的聚类分析是不全面或不彻底的。本研究为验证聚类的代表性,进行了Mantel检

测,发现 33 样株综合得分的 DIST 距离系数的空间距离和距离矩阵间全相关($r = 0.81$, $p = 1.00$),表明聚类分析能有效的表示 33 样株表型性状的相似(异)性。

因子分析使各样株有“序”或被“定量”,聚类分析使各样株有“类”或被“定性”。由“定量”到“定性”、再到“定量”,“定量”与“定性”结合,互为补充,保证了分析与综合评价的科学、客观、全面。与传统的表型选择分析评价方法相比,新的表型选择分析评价方法视角更全面,技术手段更科学、客观、便捷,结果更直观。

参考文献:

- [1] Tomilin C D. The Pesticide Manual. 12nd ed. [M]. UK: Surrey, British Crop Protection Council, 2000.
- [2] National R C. Neem-A Tree for Solving Global Problems [M]. Washington D C: National Academy Press, 1992.
- [3] 彭兴民,吴疆翀,郑益兴,等. 印楝属(*Azadirachta* A. Juss.) 植物分类及分布的研究现状[J]. 植物遗传资源学报, 2012, 13(4): 583 - 588.
- [4] Kaushik N, Singh B G, Tomar U K, et al. Regional and habitat variability in azadirachtin content of Indian neem (*Azadirachta indica* A. Jusieu) [J]. Current Science, 2007, 92(10): 1400 - 1406.
- [5] 吴疆翀. 印楝繁育系统的研究[D]. 北京: 中国林业科学研究院, 2007.
- [6] 彭兴民,吴疆翀,郑益兴,等. 云南引种印楝实生种群的表型变异[J]. 植物生态学报, 2012, 36(6): 560 - 571.
- [7] 彭兴民,吴疆翀,郑益兴,等. 云南引种印楝实生栽培种群表型的地理变异[J]. 林业科学研究, 2013, 26(4): 399 - 405.
- [8] 彭兴民,吴疆翀,程金焕,等. 印楝农药原料林优树选择方法与标准[J]. 福建林学院学报, 2010, 30(3): 265 - 269.
- [9] Sidhu O P, Kumar V, Behl H M. Variability in Neem (*Azadirachta indica*) with respect to azadirachtin content [J]. J Agri Food Chem, 2003, 51(4): 910 - 915.
- [10] Tomar U K, Singh G, Kaushik N. Screening *Azadirachta indica* tree for enhancing azadirachtin and oil contents in dry areas of Gujarat [J]. Journal of Forestry Research, 2011, 22(2): 217 - 224.
- [11] Quinn G P, Keough M J. 生物实验设计与数据分析 [M]. 蒋志刚,李春旺,曾岩,译. 北京: 高等教育出版社, 2003.
- [12] Tan Y F, Corke H. Factor analysis of physicochemical properties of 63 rice varieties [J]. Journal of the Science of Food and Agriculture, 2002, 82(7): 745 - 752.
- [13] Toker C, Cagircan M I. The use of phenotypic correlations and factor analysis in determining characters for grain yield selection in chickpea (*Cicer arietinum* L.) [J]. Hereditas, 2004, 140(3): 226 - 228.
- [14] Kurtanjek Z, Horvat D, Magdic D, et al. Factor analysis and modelling for rapid quality assessment of croatian wheat cultivars with different gluten characteristics [J]. Food Technology and Biotechnology, 2008, 46(3): 270 - 277.
- [15] 敖妍. 因子分析法在文冠果优良单株选择中的应用 [J]. 华南农业大学学报, 2009, 30(4): 70 - 73.
- [16] 马庆华,李永红,梁丽松,等. 冬枣优良单株果实品质的因子分析与综合评价 [J]. 中国农业科学, 2010, 43(12): 2491 - 2499.
- [17] 康立宁,田志刚,刘香英,等. 大豆组织蛋白产品品质的因子分析和综合评价研究 [J]. 吉林农业科学, 2012, 37(2): 45 - 48.
- [18] 宗乾收,林军,武永昆,等. 印楝种仁中印楝素含量的快速液相色谱分析 [J]. 农药, 2003, 42(4): 23 - 24.
- [19] 阮敬. SAS 统计分析从入门到精通 [M]. 北京: 人民邮电出版社出版, 2009.
- [20] 汤飞宇,程锦,黄文新,等. 高品质陆地棉数量性状的因子分析及品种聚类分析 [J]. 江西农业大学学报, 2008, 30(6): 1014 - 1018.
- [21] 柳伟伟,胡良平. 主成分分析与探索性因子分析的 SAS 软件实现 [J]. 中西医结合学报, 2010, 8(6): 589 - 593.
- [22] 胡建军. 烟叶质量评价方法优选与实证研究 [D]. 长沙: 湖南农业大学. 2009.
- [23] 徐克学. 生物数学 [M]. 北京: 科学出版社, 1999.