

# 文冠果数量性状的主成分分析及聚类分析研究

郭军战, 张 敏\*, 费昭雪, 张玲玲, 姜晓丹, 杨 梅

(西北农林科技大学 林学院, 陕西 杨陵 712100)

**摘 要:**应用 DPS 统计软件,对文冠果 12 个数量性状进行了主成分分析和聚类分析。结果表明:前 5 个主成分的累积贡献率达 70.26%,根据各性状对主成分影响作用大小及主要指标组合,将 12 个生物学性状分为 5 类,分别称之为树形因子、叶形因子、果形因子、花序因子和综合因子。根据聚类分析结果,将 12 个数量性状聚为 6 大类。2 种分析共同揭示了这几类因子对文冠果的作用。

**关键词:**文冠果;主成分分析;聚类分析;数量性状

**中图分类号:**S718.46      **文献标志码:**A      **文章编号:**1001-7461(2012)02-0066-04

## Principal Component Analysis and Cluster Analysis of Quantitative Characters of *Xanthoceras sorbifolia*

GUO Jun-zhan, ZHANG Min\*, FEI Zhao-xue, ZHANG Ling-ling, JIANG Xiao-dan, YANG Mei

(College of Forestry, Northwest A&F University, Yangling, Shaanxi 712100, China)

**Abstract:** Principal component analysis and cluster analysis were conducted on 12 quantitative characters of *Xanthoceras sorbifolia* by using the statistical software DPS. The results indicated that the additive contributing rates of the first five principal components amounted to 70.26%. According to the effect of each character on principal component and the composition of main indices, 12 biological characters could be divided into 5 groups: tree form, leaf form, fruit shape, inflorescence, and comprehensive factors. According to the result in cluster analysis, the 12 quantitative characters were divided into 6 groups. The two analyses revealed the influences of these factors on the biology of *X. sorbifolia*.

**Key words:** *Xanthoceras sorbifolia*; principal component analysis; cluster analysis; quantitative character

文冠果(*Xanthoceras sorbifolia*)属无患子科(Sapindaceae)文冠果属(*Xanthoceras*),1 属 1 种,为落叶乔木或灌木<sup>[1]</sup>,历来主要以野生状态分布在我国北方,因此其表型性状间差异较大,具有广泛的遗传变异基础。文冠果是我国特有的木本油料植物,耐干旱、瘠薄,有较强适应性和抗逆性,栽植成活率高,生长成林快,是绿化荒山,防止水土流失,改造生态环境的优良树种<sup>[2-4]</sup>。近年来,文冠果被确定为生物能源树种,其食用、药用、工业加工等综合经济价值颇高<sup>[5-9]</sup>,越来越得到人们的重视。利用主成分分析和聚类分析对文冠果生物学性状的研究迄今未见报道。一般在处理多元样本数据时,样本之间常

存在着相关关系,如文冠果的地径与冠径和树高、花芽总数和结果数、叶长和叶宽、果长和果宽等。在这些复杂的关系中,为文冠果生长结果所提供的信息大部分是重叠的,如果仅凭经验直观地判断性状间的相互关系,往往发生偏差。而主成分分析则是将多个性状指标转化为少数指标的一种统计方法,能从众多性状中找出主要的综合性状,尽可能多地反映原始数据的信息,简化数据,揭示变量之间的关系<sup>[10-14]</sup>。本研究通过对影响文冠果生长发育的几个少数主要因素进行分析及对表型性状进行分类,为今后从性状上对文冠果进行初步分类及优良单株选择上提供了数量遗传依据。

收稿日期:2011-09-01 修回日期:2011-11-27

基金项目:陕西省攻关项目“能源植物文冠果良种选育”研究(K331021003)。

作者简介:郭军战,男,副教授,硕士生导师,主要从事林木遗传育种的教学与研究。E-mail:guojunzhan@163.com

\* 通讯作者:张敏,女,在读硕士,研究方向:林木遗传育种。E-mail: zhy273@163.com

# 1 材料与方法

## 1.1 试验地概况

试验地设在西北农林科技大学林学院苗圃。该地位于关中平原中部地区<sup>[15]</sup>,地理位置处于 34°19′19″N、108°05′30″E,海拔高度 464.7 m,属暖温带半湿润气候区。年均温 10.7~13.7℃,最热月温度 24~27℃,最冷月温度-0.7~3.0℃,年极端最低气温平均-11.1~14.7℃,年降水量 500~700 mm,年均干燥度 1.09~1.67,全年≥10℃积温 3 400~4 600℃,有效生长季 152~191 d。初霜期 10 月下旬到 11 月初,晚霜期 4 月中上旬,无霜期 184~216 d。全年日照 1 900~2 500 h。土壤为垆土,土层深厚,土质适中,土壤结构良好。

## 1.2 材料收集

供试材料为 3 年生文冠果实生播种苗,郁闭度 0.5~0.7,株行距 2 m×2 m。林分健康状况良好,共观测结实单株 150 株。分别测定 12 个数量性状:地径  $X_1$ 、枝下高  $X_2$ 、冠径  $X_3$ 、树高  $X_4$ 、叶长  $X_5$ 、叶宽  $X_6$ 、果长  $X_7$ 、果宽  $X_8$ 、花芽总数  $X_9$ 、花序长  $X_{10}$ 、小花花柄长  $X_{11}$ 、结果数  $X_{12}$ 。

其中叶片性状的测量为在树冠的上、中、下 3 个部位分别随机采摘 15 片成熟叶片,测量叶片长宽;由于每株树结果量不大,将所结果实全部用游标卡尺分别测出果实的纵径和横径;当全株达到盛花期时,在南北方向各随机测量 20 个花序的长度,于每个花序中部测量 4 朵小花的花柄长,以上各指标均取平均值。

## 1.3 统计方法

试验数据采用 DPS 软件进行统计分析。①主成分分析:设  $M$  维样本数据矩阵 $[X]$ ,经标准化即:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{S_j}, \text{ 其中 } \bar{x}_j = \sum_{i=1}^n X_{ij}/n, S_j =$$

$$\sqrt{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2 / (n-1)}。 \text{ 样本矩阵的相关系数矩}$$

$$\text{阵为}[R], R_{ij} = R_{ji} = \frac{X_{ij}}{\sqrt{S_{ij} S_{ji}}} (i, j = 1, 2, \cdots, m), \text{ 从而}$$

形成特征方程: $Rb = \lambda b$ ;用雅可比法求相关系数矩阵  $R$  的全部非负特征根  $\lambda_1 > \lambda_2 > \cdots > \lambda_n \geq 0$  及相应的特征向量以及对应的特征向量  $b_1, b_2, \cdots$  的顺序,求其主成分  $y_h = b_{1i}x_1 + b_{2i}x_2 + \cdots + b_{ni}x_m, (h < m)$  以及贡献率、累积贡献率、因子负荷量<sup>[10]</sup>。②聚类分析:对原始数据标准化处理后,应用欧氏距离  $d_{ij} =$

$$\sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}, \text{ 类平均法 (UPGMA) 进行 } R \text{ 型聚类分析。}$$

其中  $n$  为样本个数,  $m$  为性状个数,  $x_{ij}$  为第  $i$  个样本的第  $j$  个性状的原始数据;  $x_{ij}^*$  为  $x_{ji}$  经过变换后的数据;  $\bar{x}_j$  为第  $j$  个性状数据的平均值;  $S_j$  为第  $i$  个品种的第  $j$  个性状的标准差;  $d_{ij}$  为第  $i$  个分类单位与第  $j$  个分类单位的距离系数。

# 2 结果与分析

## 2.1 主成分分析

各因子之间的相关系数用相关矩阵来表示。从表 1 可以看出,相关关系明显的因子有:地径  $X_1$  与冠径  $X_3$  和树高  $X_4$ ,花序长  $X_{10}$  和小花花柄长  $X_{11}$ ,花芽总数  $X_9$  和结果数  $X_{12}$ ,叶长  $X_5$  和叶宽  $X_6$ ,果长  $X_7$  和果宽  $X_8$ 。这些性状之间的相关程度达到了统计学水准的  $r_{0.05}$  或  $r_{0.01}$ 。

从表 2 可以看出,前 5 个主成分的特征根均大于 1,累积贡献率达 70% 以上,代表了原始数据的大部分信息。

进一步分析前 5 个主成分各性状的特征向量值及其主要性状指标组合(表 3)。

由于主成分是 12 个数量性状的线性组合函数。根据计算相关矩阵的特征向量见表 3 可给出主成分的函数式为:

$$Y_1 = 0.382\ 3X_1 + 0.133\ 9X_2 + 0.351\ 0X_3 + 0.394\ 6X_4 + 0.367\ 1X_5 + 0.271\ 1X_6 + 0.254\ 7X_7 + 0.157\ 0X_8 + 0.241\ 9X_9 + 0.277\ 7X_{10} + 0.338\ 6X_{11} + 0.298\ 9X_{12}$$

$$Y_2 = -0.340\ 9X_1 - 0.004\ 7X_2 - 0.410\ 2X_3 - 0.243\ 7X_4 + 0.437\ 0X_5 + 0.483\ 1X_6 + 0.173\ 4X_7 + 0.068\ 4X_8 - 0.379\ 6X_9 + 0.252\ 6X_{10} + 0.259\ 9X_{11} - 0.212\ 6X_{12}$$

.....

由以上函数式和表 3 可以看出,第 1 主成分中,树高  $X_4$  和地径  $X_1$  的系数最大,可见第 1 主成分表现为树形因子的影响。第 2 主成分中,叶宽  $X_6$  和叶长  $X_5$  的系数最大,所以第 2 主成分表现为叶形因子的影响。第 3 主成分中,果宽  $X_8$  和果长  $X_7$  的系数最大,所以第 3 主成分表现为果形因子的影响。第 4 主成分中,花序长  $X_{10}$ 、小花花柄长  $X_{11}$  的系数最大,所以第 4 主成分表现为花序因子的影响。第 5 主成分中,枝下高  $X_2$ 、果长  $X_7$ 、花序长  $X_{10}$  的系数最大,所以第 5 主成分表现为树形、果形、花序因子的综合影响。

## 2.2 聚类分析

从图 1 可以看出,当距离系数阈值为 7.96 时,12 个表型性状可分为 6 类。

表 1 各性状相关矩阵  
Table 1 Correlation coefficient of each character

性状	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>
X <sub>1</sub>	1.000 0					
X <sub>2</sub>	0.232 7	1.000 0				
X <sub>3</sub>	0.291 4*	−0.250 6	1.000 0			
X <sub>4</sub>	0.325 4*	0.190 9	0.253 8	1.000 0		
X <sub>5</sub>	0.119 7	−0.143 4	−0.151 4	0.266 5	1.000 0	
X <sub>6</sub>	−0.046 0	0.124 1	0.025 1	−0.176 0	0.845 2**	1.000 0
X <sub>7</sub>	−0.004 7	−0.000 3	0.138 5	−0.066 6	0.210 9	−0.034 5
X <sub>8</sub>	0.094 3	0.150 3	0.084 5	−0.059 7	−0.151 6	−0.011 0
X <sub>9</sub>	0.136 3	0.214 6	0.133 3	0.039 6	0.000 8	0.021 9
X <sub>10</sub>	−0.059 6	0.177 0	0.004 0	0.033 7	0.151 5	−0.106 6
X <sub>11</sub>	−0.088 5	0.003 9	0.072 8	0.106 9	0.031 7	0.151 7
X <sub>12</sub>	−0.012 2	−0.233 7	0.206 3	−0.045 4	0.171 5	−0.187 3

性状	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>
X <sub>7</sub>	1.000 0					
X <sub>8</sub>	0.527 6**	1.000 0				
X <sub>9</sub>	0.204 3	−0.209 2	1.000 0			
X <sub>10</sub>	0.142 3	0.067 1	−0.273 5	1.000 0		
X <sub>11</sub>	0.033 7	0.104 6	−0.167 9	0.328 0*	1.000 0	
X <sub>12</sub>	−0.304 0	0.081 1	0.478 7**	0.212 3	0.269 7	1.000 0

注：\* 表示差异显著；\*\* 表示差异极显著。

表 2 主成分的特征根、贡献率、累积贡献率  
Table 2 Characteristic root contribute proportios and additive contribute proportios of principal components

主成分	特征根	贡献率/%	累积贡献率/%
1	2.694 5	23.319 3	22.172 8
2	1.974 2	16.068 1	41.063 9
3	1.703 5	14.196 2	54.966 4
4	1.653 5	13.778 8	70.264 8
5	1.002 7	8.857 0	72.996 5
6	0.858 3	7.243 2	85.743 3
...	...	...	...

表 3 前 5 个主成分相关矩阵的特征向量  
Table 3 Eigenvectors of correlation matrix of the first five principal components (PC)

性状	第 1 主成分	第 2 主成分	第 3 主成分	第 4 主成分	第 5 主成分
X <sub>1</sub>	0.382 3	−0.340 9	0.134 8	−0.267 1	0.082 0
X <sub>2</sub>	0.133 9	−0.004 7	0.341 3	−0.375 6	0.584 3
X <sub>3</sub>	0.351 0	−0.410 2	0.032 3	0.200 5	−0.287 3
X <sub>4</sub>	0.394 6	−0.243 7	−0.002 2	−0.179 7	0.298 8
X <sub>5</sub>	0.367 1	0.437 0	−0.284 4	−0.254 9	−0.150 3
X <sub>6</sub>	0.271 1	0.483 1	−0.233 1	−0.333 2	−0.186 0
X <sub>7</sub>	0.254 7	0.173 4	0.516 9	−0.025 6	−0.457 6
X <sub>8</sub>	0.157 0	0.068 4	0.602 3	0.238 9	−0.195 4
X <sub>9</sub>	0.241 9	−0.379 6	−0.207 0	−0.218 7	−0.231 9
X <sub>10</sub>	0.277 7	0.252 6	0.073 5	0.408 3	0.392 3
X <sub>11</sub>	0.338 6	0.259 9	−0.098 8	0.407 4	0.172 5
X <sub>12</sub>	0.298 9	−0.212 6	−0.386 1	0.333 1	0.014 2

第 1 类由地径 X<sub>1</sub>、树高 X<sub>4</sub>、冠径 X<sub>3</sub> 聚为一类，称之为树形因子；

第 2 类由花芽总数 X<sub>9</sub>、结果数 X<sub>12</sub> 聚为一类，称之为坐果因子；

第 3 类由枝下高 X<sub>2</sub> 为一类；

第 4 类由果长 X<sub>7</sub>、果宽 X<sub>8</sub> 聚为一类，称之为果形因子；

第 5 类由叶长 X<sub>5</sub>、叶宽 X<sub>6</sub> 聚为一类，称之为叶形因子；

第 6 类由花序长 X<sub>10</sub>、小花花柄长 X<sub>11</sub> 聚为一类，称之为花序因子。

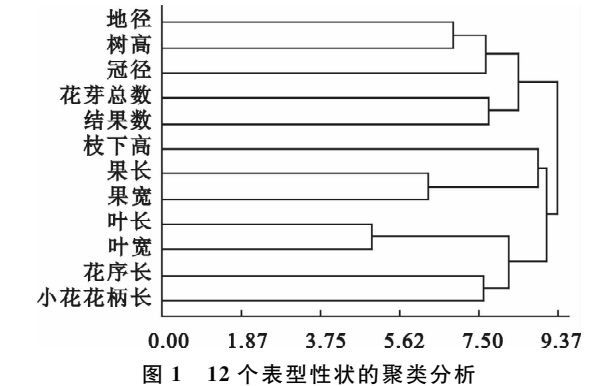


Fig. 1 Cluster analysis of twelve biological characters

3 结论与讨论

应用多元统计中的主成分分析法，探讨了文冠果数量性状间的相互关系及其分类。12 个观测性状间地径与冠径和树高、花序长和小花花柄长、花芽总数和结果数、叶长和叶宽、果长和果宽相关性最

大,其他性状间相关系数较小。将文冠果主要生物学性状指标转化为较少的几个主成分。其中前 5 个主成分的累积贡献率达 70%以上,其所表达的综合信息可以用来表达全部性状的信息。通过 *R* 型聚类分析法将 12 个数量性状分为 6 类:树形因子、叶形因子、果形因子、花序因子、坐果因子和枝下高。

文冠果的生物学性状较多,且相互间存在错综复杂的关系,致使它们提供的信息出现重叠,不易得出简明的规律。采用主成分分析和 *R* 型聚类分析简化了评价指标,从而能够比较容易和清晰地解释其间的相关关系。2 种分析共同揭示了影响文冠果生长的主要因子组成。为今后进行文冠果分类研究提供新的理论依据。本研究仅探讨了文冠果 12 个数量性状间的关系,若采用更多性状因子进行分析,其结果如何,尚待进一步研究。

参考文献:

[1] 徐东翔. 文冠果[M]. 内蒙古:内蒙古人民出版社,1977:1.  
XU D X. *Xanthoceras sorbifolia* Bunge[M]. Inner Mongolia: Inner Mongolia People Press,1977:1. (in Chinese)

[2] 马启慧. 能源树种文冠果的研究现状与发展前景[J]. 北方园艺,2007,2(8):77-78.  
MA Q H. Research present situation and development prospect of energy tree species *Xanthoceras sorbifolia* Bunge[J]. Northern Horticulture,2007,2(8):77-78. (in Chinese)

[3] 董云岚. 文冠果适生条件调查[J]. 河南林业科技,1985,3(1):35-38.  
DONG Y L. Investigation of *Xanthoceras sorbifolia* Bunge optimal growth condition[J]. Journal of Henan Forestry Science and Technology, 1985,3(1):35-38. (in Chinese)

[4] 白彤. 文冠果利用与栽培[M]. 内蒙古:内蒙古人民出版社,2009:15-16.  
BAI T. Utilization and cultivation of *Xanthoceras sorbifolia* Bunge[M]. Inner Mongolia: Inner Mongolia People Press, 2009:15-16. (in Chinese)

[5] 霍力彬,侯立群. 中国文冠果资源研究开发与实践[M]. 北京:中国林业出版社,2007:33-39.  
HUO L B, HOU L Q. Development and Practice of Research on China *Xanthoceras sorbifolia* Bunge Resources[M]. Beijing: China Forestry Press,2007:33-39. (in Chinese)

[6] 特格喜,乌力吉. 文冠果治疗遗尿症 120 例临床观察[J]. 中国民族医药杂志,1997,3(1):19.  
TE G X, WU L J. Clinical observation on 120 cases of *Xanthoceras sorbifolia* Bunge to treat enuresis[J]. Journal of Medicine & Pharmacy of Chinese Minorities, 1997, 3(1):19. (in Chinese)

[7] YAKUGAKU Z. Inhibitory effects of some traditional medicines on proliferation of HIV-1 and its protease[J]. Journal of The Pharmaceutical Society of Japan, 2004, 124(8):519-526.

[8] LI Z L, LI X, YANG B Z. Two new triterpenes from the husks of *Xanthoceras sorbifolia* [J]. Planta Medica, 2005,71(11):1068-1071.

[9] 尚宏芹. 文冠果综合利用研究进展[J]. 生物学教学,2010,35(7):16-17.  
SHANG H Q. Comprehensive utilization and research progress of *Xanthoceras sorbifolia* Bunge[J]. Biology Teaching, 2010,35(7):16-17. (in Chinese)

[10] 唐启义,冯明光. DPS 数据处理系统-实验设计、统计分析 & 数据挖掘[M]. 北京:科学出版社,2007:696-722.  
TANG Q Y, FENG M G. DPS Data Processing System-experimental Design, Statistical Analysis and Data Mining[M]. Beijing: Science Press,2007:696-722. (in Chinese)

[11] ANAERSON T W. Introduction to multivariate statistical analysis[M]. New York:Wiley,1967.

[12] GAUCH H G J. Multivariate Analysis in Community Ecology[M]. Cambrige: Cambridge University Press,1982.

[13] 唐守正. 多元统计分析方法[M]. 北京:中国林业出版社,1986:583-620.  
TANG S Z. Multivariate Statistical Analysis Method[M]. Beijing: China Forestry Press,1986:583-620. (in Chinese)

[14] 王学仁,王松维. 实用多元统计分析[M]. 上海:上海科技出版社,1990,570-609.  
WANG X R, WANG S W. Practical Multivariate Statistical Analysis[M]. Shanghai: Shanghai Science and Technology Press,1990:570-609. (in Chinese)

[15] 熊晓锐,廖允成,高茂盛. 黄土高原东南部旱作农田一年二熟种植模式水分效应初探——以杨凌为例[J]. 干旱地区农业研究,2008,26(3):44-49.  
XIONG X R, LIAO Y C, GAO M S. Study on effect of double cropping system on moisture of dry farmland in the southeast Loess Plateau——A case study of Yangling[J]. Agricultural Research in the Arid Areas, 2008,26(3):44-49. (in Chinese)