

主成分分析에 의한 在來種 옥수수의 解析

이인섭* · 박종옥

경성대학교 자연과학부

Assessment and Classification of Korean Indigenous Corn Lines by Application of Principal Component Analysis

In Sup Lee* and Jong Ok Park

College of Science, Kyungsung University, Busan 608-736, Korea

Abstract

This study was conducted to get basic information on the Korean local corn line collected from Busan City and Kyungnam Province, a total of 49 lines were selected and assessed by the principal component analysis method. In the result of principal component analysis for 7 characteristics, 67.4% and 86.3% of total variation could be appreciated by the first two and first four principal components, respectively. Contribution of characteristics to principal component was high at upper principal components and low at lower principal components. Biological meaning of principal component and plant types corresponding to the each principal component were explained clearly by the correlation coefficient between principal component and characteristics. The first principal component appeared to correspond to the size of plant and ear, and the duration of vegetative growing period. The second principal component appeared to correspond to the number of ear and tiller. But the meaning of the third and fourth principal components were not clear.

Key words – Korean local corn line, principal component analysis, characteristics

서 론

作物育種에 필요한 育種材料를 얻기 위해서는 풍부한 遺傳子源의 확보가 중요하다. 이러한 유전자원의 확보를 위해서는 아직껏 育種材料로 이용된 바가 없는 在來種을 수집하는 등의 방법도 도움이 될 것이다. 또한 확보된 유전자원은 그 특성에 따라 分類하고 整理한 후 育種目標에 맞게 이용하는 것이 필요하다. 그리고 확보된 유전자원에 대하

여 特性調査를 하고 分類를 할 때에는 충분히 效用價值가 있도록 해야 하며, 많은 系統과 形質을 대상으로 하여 客觀性을 가지도록 하여야 할 것이다.

근래에 이르러 컴퓨터가 널리 보급되고 많은 통계 프로그램이 개발·보급되면서 多變量解析法이 활발히 이용되어 왔고, 객관성을 띤 品種分類에도 적용하게 되었다. 그 중에서도 種間相關이나 品種間 距離에 의한 cluster 分析 및 主成分分析 등이 下等生物이나 高等植物의 種間 遺傳的 類緣關係를 밝히는 數理學的 分類方法으로 이용되어 왔다. 특히 主成分分析은 하등생물과 동물의 變異研究에 많이 적용되었으며, 고등식물에서 적용한 예를 보면, 옥수

*To whom all correspondence should be addressed
Tel : 051-620-4647, Fax : 051-627-4115
E-mail : yslee@star.ks.ac.kr

수 菲集系統의 分類[8, 9, 10], 옥수수 收量試驗에서 調査形質의 選擇[6], 果樹의 體刑變異解析[12], 벼 三染色體의 分類[11], 밀의 自然變異解析을 위한 形質의 集約[2], 벼의 雜種系統에 대한 遺傳的 類緣關係[4], 油株의 分類[3] 등에 이를 적용한 보고가 있다.

본 연구에서는 부산·경남지역에서 수집한 재래종 옥수수 중에서 선발한 49 계통의 특성을 조사하고, 그 特性에 따라 재래종 옥수수를 解析하고 分類하여 앞으로의 옥수수 育種에 이용할 基礎資料를 얻고자 하였다.

재료 및 방법

1998 부산·경남지역에서 수집한 재래종 옥수수[7] 중에서 병충해에 강하고 비교적 환경 적응성이 높은 49 계통을 부산광역시 기장군 기장읍 교리에 1999년부터 2001년까지 재배하고 특성조사를 행하였다. 재배 방법으로는 500cm × 70cm 이랑에 系統당 15株씩 30cm 간격으로 첨파하였고, 1株 1本식으로 하였으며, 그 중에서 가능하면 病蟲害가 없는 10個體를 임의로 골라 조사하였다. 供試系統의 特性検定을 위한 조사항목은 表1에서 보는 바와 같다.

재래종 옥수수의 해석 및 계통분류를 위해 본 연구에 사용된 統計的方法은 主成分分析은 數理統計的理論은 Anderson [1], Kendal[5], Seal[12], Sokal[13] 등에 의하여 자세히 소개되었고, 국내에서는 崔와 李[3]가 유채를 분류하는데, 李와 崔[7]가 한국 재래종 옥수수를 분류하는데 소개된 바 있다.

결과

主成分의 抽出- 固有值

고유치는 각각의 주성분으로 설명할 수 있는 변수들의 분산의 총합으로서, 변수 속에 담겨진 정보(분산)가 어떤 주성분에 의하여 어느 정도 표현할 수 있는가를 말해주는 비율을 의미한다. 따라서 고유치는 각 주성분별로 모든 변수의 요인적재값을 제곱하여 더한 값으로 계산된다. Table 1에서 보는 바와 같이 각 변수들의 단위가 다르기 때문에 먼저 각 변수들을 표준화 한 후 주성분분석을 하였다.

고유치 및 전체변동(변수의 수)에 대한 주성분의 기여율은 Table 2에서 보는 바와 같으며, 제 1, 2, 3, …, n 주성분 순으로 3.341, 1.377, 0.696, … 으로 작아지는데, 제 1 주성

Table 1. Characteristics of Korean indigenous corn lines used in this study

| No. | Characteristics | Unit |
|-----|--------------------|------|
| 1 | Ear length | cm |
| 2 | Ear weight | gr |
| 3 | 100 kernel wt. | gr |
| 4 | Ear number | — |
| 5 | Days to tasselling | — |
| 6 | Tiller number | — |
| 7 | Plant height | cm |

분, 제 2 주성분, 제 3 주성분에 의하면 전체 변동[6]의 47.735%, 19.670%, 9.938%를 각각 설명할 수 있다. 따라서 제 1 주성분에서 제 4 주성분까지만 취했을 경우에도 전체 변동의 86.311%를 설명할 수가 있는 것이다.

고유치가 1보다 크다는 것은 하나의 주성분이 변수 한 개 이상의 분산을 설명해 준다는 것을 의미하므로 일반적으로 고유치가 1보다 큰 주성분만을 추출한다. 따라서 제 1 주성분과 제 2 주성분의 고유치가 1보다 크므로 본 연구에서는 제 1 주성분과 제 2 주성분만을 추출하여 논하기로 한다.

변수와 주성분과의 상관계수는 각 주성분에 대응한 고유치의 평방근에 의하여 그 크기가 결정되는데 변수와 주성분과의 관계를 각 주성분을 통하여 비교하는데 상관계수가 유용하다. Table 3에서 보는 바와 같이 상위 주성분에서 상관이 높은 형질이 많았고, 하위 주성분에서 상관이 높은 형질이 적었다.

주성분과 형질간의 상관계수의 자승한 값은 형질의 주성분에 대한 기여율을 나타낸다. 이 기여율은 추출한 n개의 주성분에 의해 설명되는 형질 X_i 의 변동의 비율을 나타내고 있다. Table 3의 오른쪽 끝에서는 제 1 주성분에서 제 4 주성분까지의 누적 기여율을 보여주고 있다. 제 4 주성분까지의 기여율은 이삭수에서 99.4%로 최고를 나타내었고, 草長에서 68.0%로서 최저로 나타났으며 전체 평균값은 86.3%이었다.

Fig. 1은 Table 3의 각 주성분의 상관계수의 값을 2차 평면(X_1, X_2) 위에 도표화한 성분도표를 나타내고 있다.

또한 Table 4에서는 주요형질과 주성분간의 상관관계를 그 값에 따라 正, 負 2계급으로 구분하고 正 또는 負에서

主成分分析에 의한 在來種 옥수수의 解析

Table 2. Eigen value obtained from principal component analysis

| Principal component | X ₁ | X ₂ | X ₃ | X ₄ |
|----------------------------|----------------|----------------|----------------|----------------|
| Eigen value | 3.341 | 1.377 | 0.696 | 0.628 |
| Contribution(%) | 47.735 | 19.670 | 9.938 | 8.968 |
| Cumulative eigen value | 3.341 | 4.718 | 5.414 | 6.042 |
| Cumulative contribution(%) | 47.735 | 67.405 | 77.343 | 86.311 |
| Principal component | X ₅ | X ₆ | X ₇ | |
| Eigen value | 0.468 | 0.429 | 0.061 | |
| Contribution(%) | 6.682 | 6.131 | 0.877 | |
| Cumulative eigen value | 6.510 | 6.939 | 7.000 | |
| Cumulative contribution(%) | 92.993 | 99.123 | 100.000 | |

Table 3. Correlation coefficient between characteristics and principal component, and cumulative contribution of characteristics

| Characteristics | Principal Component | | | | Cumulative contribution |
|--------------------|---------------------|----------------|----------------|----------------|-------------------------|
| | X ₁ | X ₂ | X ₃ | X ₄ | |
| Ear length | 0.870** | 0.080 | 0.350 | 0.133 | 0.903 |
| Ear weight | 0.910** | 0.058 | 0.221 | 0.132 | 0.898 |
| 100 kernel wt. | 0.733** | -0.271 | -0.336 | -0.317 | 0.824 |
| Ear No. | 0.400 | 0.629** | -0.557 | 0.358 | 0.994 |
| Days to tasselling | -0.714** | 0.358 | 0.232 | 0.302 | 0.783 |
| Tiller No. | -0.227 | 0.796** | 0.074 | -0.518 | 0.959 |
| Plant height | 0.705** | 0.369 | 0.206 | -0.067 | 0.680 |

**Significant difference at 1% level.

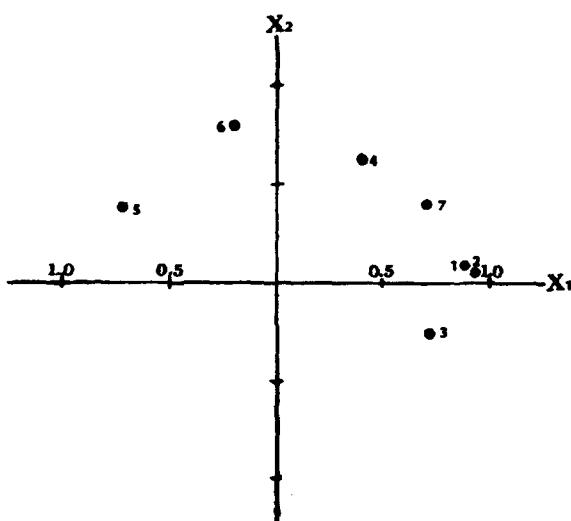


Fig. 1. Scatter diagram of characteristics of Korean local corn lines with standardized value in the X_1 - X_2 plane.

Number's stand for characteristics (1; ear length, 2; ear weight, 3; 100 kernel weight, 4; ear number, 5; days to tasselling, 6; tiller number, 7; plant height)

Table 4. Classification of characteristics by the degree of contribution to the first two principal components

| Principal component | Class | Corresponding characteristics |
|---------------------|-------|--|
| X_1 | + | Ear weight, Ear length, 100 kernel wt., Plant height |
| | - | Days to tasselling |
| X_2 | + | Tiller No., Ear No. |
| | - | None |

有意(5% 水準)한 形質만을 나타내었다. 계급 구분은 Table 3에 나타난 주성분과 형질과의 상관관계를 이용하였다.

Fig. 1과 Table 4에서 보면 제 1 주성분의 경우 係數가
正인 形質은 그 값이 큰 순으로 이삭重, 이삭길이, 100粒重,
草長 등이었고, 負의 형질은 出穗期까지의 日數이었다. 이
주성분에서 주성분 점수가 크기 위해서는 이삭길이, 이삭
重, 100粒重, 草長 등의 값이 커야하며, 출수기까지의 일수
작은 값을 것을 필요로 한다. 즉 이산의 깊이, 草長 등이 크

고 이삭의 무게가 무거울수록 그리고 출수기까지의 일수 등의 값이 작을수록 제 1 주성분의 점수는 큰 값을 가지게 된다. 이와 같이 제 1 주성분에 의한 계통의 변이 방향은 이삭길이, 무게, 식물체의 크기가 大型化하는 방향과 출수기까지의 날짜수가 小型化하는 방향에 있다고 할 수 있다. 이상으로부터 제 1 주성분은 식물체의 크기(영양기관 및 생식기관)와 출수기에 관련된 주성분이라 할 수 있고, 이 주성분에 의하여 短稈小穗의 晚生種과 長稈大穗의 早生種의 구별이 가능하였다.

이러한 결과는 출수기까지의 일수에서 李[6]가 보고한 바로는 제 2 주성분의 正의 유의한 값이었는데 본 연구에서는 제 1 주성분의 負의 有意한 값을 가지게 되어 차이가 있었는데, 이것은 주성분분석에 이용된 形質의 차이에 기인하는 것으로 생각된다.

제 3 주성분과 제 4 주성분에서는 유의한 형질이 없었다.

主成分의 點數

49 系統의 主成分 點數는 Table 5에서 보는 바와 같고, 제 1 주성분에 의한 系統分布図는 Fig. 2에서 보는 바와 같다. 주성분은 이론적으로 直交하기 때문에 제 2 주성분까지를 추출할 경우 두 성분을 座標軸으로 하는 2次元 平面上

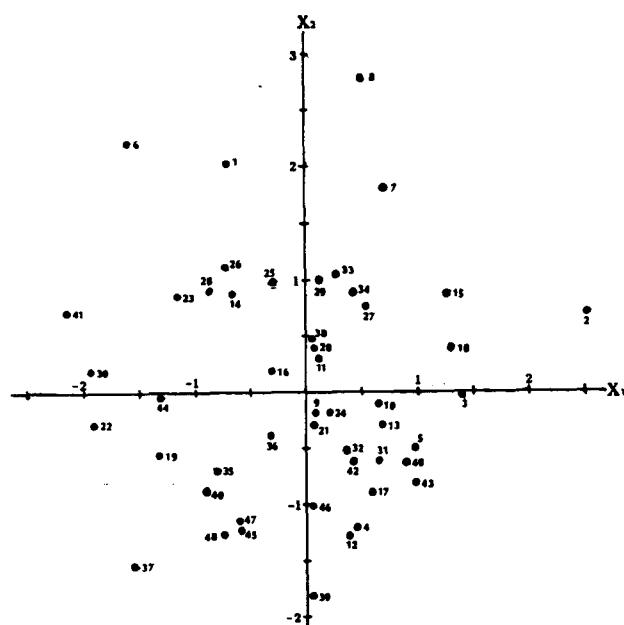


Fig. 2. Scatter diagram of Korean local corn lines in the X_1 - X_2 plane.

Table 5. Principal component score of Korean local corn lines

| Line No. | Principal component | | | |
|----------|---------------------|--------|--------|--------|
| | X_1 | X_2 | X_3 | X_4 |
| 1 | -0.694 | 2.021 | -0.439 | 0.056 |
| 2 | 2.526 | 0.703 | -0.859 | 2.116 |
| 3 | 1.446 | -0.001 | -1.349 | 0.449 |
| 4 | 0.415 | -1.160 | -0.830 | -0.210 |
| 5 | 1.073 | -0.504 | 0.961 | 0.615 |
| 6 | -1.596 | 2.172 | 0.382 | -2.053 |
| 7 | 1.230 | 1.837 | 0.755 | 0.297 |
| 8 | 0.509 | 2.772 | -0.305 | 1.768 |
| 9 | 0.095 | -0.282 | 0.449 | 0.765 |
| 10 | 0.653 | -0.092 | 0.778 | -0.103 |
| 11 | 0.128 | 0.328 | -0.149 | -1.515 |
| 12 | 0.402 | -1.307 | -1.256 | -0.612 |
| 13 | 0.727 | -0.283 | -1.815 | 0.516 |
| 14 | -0.656 | 0.874 | -2.349 | -0.766 |
| 15 | 1.285 | 0.816 | -0.178 | -0.042 |
| 16 | -0.300 | 0.202 | -0.735 | -1.705 |
| 17 | 0.646 | -0.921 | 1.221 | -1.616 |
| 18 | 1.326 | 0.456 | 0.833 | -1.478 |
| 19 | -1.273 | -0.579 | -0.166 | 1.600 |
| 20 | 0.105 | 0.373 | -0.074 | -1.423 |
| 21 | 0.068 | -0.279 | 0.178 | -0.166 |
| 22 | -1.904 | -0.373 | 0.999 | 0.322 |
| 23 | -1.215 | 0.850 | 0.073 | -0.364 |
| 24 | 0.233 | -0.181 | -1.322 | -2.399 |
| 25 | -0.377 | 0.981 | 0.770 | -0.594 |
| 26 | -0.756 | 1.144 | -1.079 | 0.585 |
| 27 | 0.518 | 0.747 | -0.253 | 0.931 |
| 28 | 0.931 | -0.920 | 2.852 | 0.404 |
| 29 | 1.032 | 0.127 | 1.417 | 0.042 |
| 30 | -1.938 | 0.187 | 0.496 | 2.400 |
| 31 | 0.678 | -0.620 | 0.027 | -0.082 |
| 32 | 0.367 | -0.560 | -0.101 | -0.044 |
| 33 | 0.277 | 1.065 | 1.712 | 0.017 |
| 34 | 0.430 | 0.792 | 0.803 | -0.827 |
| 35 | -0.776 | -0.716 | -0.678 | 0.496 |
| 36 | -0.333 | -0.350 | -0.088 | 0.182 |
| 37 | -1.510 | -1.550 | 0.457 | 0.518 |
| 38 | 0.043 | 0.491 | -1.047 | 0.602 |
| 39 | 0.010 | -1.815 | -0.066 | -1.237 |
| 40 | -0.886 | -0.873 | 1.618 | 0.738 |
| 41 | -2.147 | 0.701 | -0.089 | 0.306 |
| 42 | 0.444 | -0.618 | 0.081 | 0.246 |
| 43 | 0.996 | -0.822 | 0.120 | 0.301 |
| 44 | -1.440 | 0.023 | 1.429 | -0.688 |
| 45 | -0.550 | -1.247 | -1.049 | 0.466 |
| 46 | 0.085 | -1.028 | -0.084 | 0.711 |
| 47 | -0.584 | -1.173 | -1.514 | -0.140 |
| 48 | -0.656 | -0.747 | -0.584 | 0.616 |
| 49 | 0.908 | -0.657 | 0.051 | -0.005 |

에 49개의 계통을 나타내는 점이 분포하고 있고 점의 위치는 그 주성분 점수에 의해 주어진다. 이 평면상의 2점의 거리를 계통간 거리로 정의하면, 다수의 형질에 대해 측정한 계통의 종합적인 類緣性의 정도는 이 계통간 거리라고 하는 주성분분석에 기초하여 1차원의 수치로 표현할 수 있다. 계통간 거리가 작을수록 계통간의 유연성이 높고, 계통간 거리가 클수록 유연성은 낮게 된다.

고 칠

Mochizuki[8, 9]는 日本 在來種 옥수수를 전국적으로 落集하고 分類하였는데 育種材料의 탐색을 위한 계통의 분류는 主成分分析과 그것에 기초한 系統間 距離의 적용이 有效하다고 하였으며, 李[6]는 韓國 在來種 옥수수를 전국적으로 수집하고 이를 主成分分析을 이용하여 분류하고 解析하였는데, 12가지의 特性 중 제 4 주성분까지를 가지고 전체 변동의 67.09%가 설명될 수 있음을 알았다. 주성분분석의 目的을 생각할 때 전체 변동의 대부분(80% 이상)을 설명할 수 있는 少數의 主成分이抽出되는 것이 必要하다고 할 때[8, 9] 이는 이를 特성들을 가지고 주성분분석을 하기에 다소 미흡한 점이 있었다고 할 수 있다.

또한 주성분분석을 통한 재래종 옥수수의 解析이 育種材料의 탐색이라는 점에 목적을 두고 있기 때문에, 多數의 测定된 特性 모두를 주성분분석에 利用하는 것도 좋겠으나 目的에 맞는 特性을 어느 정도 選擇해서 주성분분석을 적용하는 편이 더 바람직하다고 생각된다. 다시 말해서 特성을 많이 取하게 되면 重要的한 特性的 비중이 상대적으로 낮아지게 되어 解析의 의미가 약화되는 것이다.

본 연구에서는 제 4 주성분까지의 기여율이 86.3%나 되어 상당히 높은 것으로 나타났으며, 제 1 주성분과 제 2 주성분만을 가지고서도 전체 변동의 67.4%를 설명할 수 있는 수준이었고, 또한 주성분분석에 이용된 7가지 特性 모두가 제 1 주성분 혹은 제 2 주성분에서 1% 水準의 有意性을 나타내어 제 1 주성분과 제 2 주성분만을 가지고서도 주성분분석이 가능한 수준이었다.

요 약

育種材料를 얻기 위하여 부산·경남지역에서 수집된 재

래종 옥수수 49 계통을 선발하여 본 실험을 실시하였다. 본 시료는 主成分分析을 이용하여 재래종 옥수수를 解析하고 系統分類를 실시하였던 바 다음과 같은 결과를 얻었다. 7 個의 形質을 이용하여 실시한 주성분분석에서는 제 4 주성분까지를 가지고 전체 변동의 86.3%를 설명할 수 있었고, 제 2 주성분까지는 전체 변동의 67.4%를 설명할 수 있었다. 주성분에 대한 형질들의 기여율은 형질에 따라 달랐고 상위 주성분에서 커으며 하위 주성분에서 작았다. 주성분과 형질과의 상관계수는 주성분의 생물학적 의의와 주성분에 대응한 식물체의 型을 명확히 하였는데 제 1 주성분은 식물체의 크기 및 생장기간에 관련된 주성분이었고, 제 2 주성분은 이삭수와 분蘖수에 관련된 주성분이었다. 제 3 주성분과 제 4 주성분에서는 형질간에는 유의성이 인정되지 않았다.

감사의 글

이 연구는 2001년도 경성대학교 연구비의 지원을 받아 수행되었습니다.

참 고 문 헌

- Anderson, T. W. 1959. An introduction to multivariate statistical analysis. John Wiley & Sons, Inc., New York.
- Chebib, F. S., S. B. Helgason and P. J. Kaltsikes. Effect of variation in plant spacing, seed size and genotype on plant-to-plant variability in wheat. Z. Pflanzenzüchtig. 69, 301-332.
- Choi, H. C. and J. I. Lee. 1979. Classification of rape seed cultivars by the principal component analysis and culster analysis. Korean J. Breeding. 11(3), 175-195.
- Fujimaki, H. 1978. Genetical studies on improvement of backcross technique in rice breeding. J. Cent. Agr. Exp. Sta. 27, 187-246.
- Kendall, M. G. 1957. A course in multivariate analysis. Charles Griffin & Company Ltd. London. p. 185.
- Lee, I. S. and B. H. Choe. 1982. Assessment and classification of Korean local corn lines by application of principal component analysis. Korean J. Breeding. 14(3), 294-303.

7. Lee, I. S. and J. O. Park. 2001. Ear and kernel characteristics of Korean indigenous maize lines collected in Pusan and Kyungnam. *Korean J. Life Sci.* **11(2)**, 159-165.
8. Mochizuki, N. 1968. Classification of breeding materials by the principal component analysis. *Bull. Natl. Inst. Agr. Sci. D.* **19**, 85-149.
9. Mochizuki, N. and T. Okuno. 1967. Classification of maize lines and selection of breeding materials by the application of multivariate analysis. *Maize Genet. Coop. News Letter* **41**, 142-147.
10. Murakami, M. and S. Nakamura. 1976. Classification of Perilla (*P. ocimoides* L.) by application of the principal component analysis. *Japan. J. Breed.* **26**, 89-90.
11. Nakagahra, M., T. Akihama and N. Iwata. 1976. Directional change of quantitative characters in trisomic in rice plants (*Oryza sativa* L.). *Japan. J. Breed.* **26(1)**, 51-58.
12. Seal, H. L. 1964. Multivariate statistical analysis for biologists. Methuen and Co. Ltd.
13. Sokal, R. R. 1961. Distance as a measure of taxonomic similarity. *Systematic Zool.* **10**, 70-79.

(Received March 17, 2003; Accepted June 18, 2003)