Principal component analysis of early generation sorghum lines for yield-contributing traits and resistance to midge

Massaoudou Hamidou, Oumarou Souleymane, Malick Ba, Eric Danquah, Issoufou Kapran, Vernon Gracen & Kwadwo Ofori

To cite this article: Massaoudou Hamidou, Oumarou Souleymane, Malick Ba, Eric Danquah, Issoufou Kapran, Vernon Gracen & Kwadwo Ofori (2018): Principal component analysis of early generation sorghum lines for yield-contributing traits and resistance to midge, Journal of Crop Improvement, DOI: 10.1080/15427528.2018.1498423

To link to this article: https://doi.org/10.1080/15427528.2018.1498423

Published online: 26 Jul 2018.
Principal component analysis of early generation sorghum lines for yield-contributing traits and resistance to midge

Massaoudou Hamidou\textsuperscript{a,c}, Oumarou Souleymane\textsuperscript{a}, Malick Ba\textsuperscript{b}, Eric Danquah\textsuperscript{c}, Issoufou Kapran\textsuperscript{a}, Vernon Gracen\textsuperscript{c}, and Kwadwo Ofori\textsuperscript{c}

\textsuperscript{a}Departement des cultures pluviales, Institut National de la Recherche Agronomique du Niger (INRAN), Niamey, Niger; \textsuperscript{b}Entomology Department, International Crop Research Institute for the Semi-Arid Tropics (ICRISAT), Niamey, Niger; \textsuperscript{c}West Africa Centre for Crop Improvement, College of Basic and Applied Sciences, University of Ghana Legon, Accra, Ghana

\begin{abstract}
Sorghum (\textit{Sorghum bicolor} L. Moench) is a staple food crop in Niger. Its production is constrained by sorghum midge (\textit{Stenodiplosis sorghicola} Coquillett 1898) and by the use of low-yielding, local sorghum varieties. To improve sorghum productivity, farmers need high-yielding and midge-resistant sorghum cultivars. The objective of this research is to evaluate genetic diversity among sorghum inbred lines for yield, yield-contributing traits, and resistance to midge. Two-hundred and eighty-two lines (280 recombinant inbred lines plus two checks) were evaluated at two experimental sites in Niger. An alpha (0.1) lattice design with two replications was used to conduct the trial. Principal component analysis was used to determine the relationship between grain yield and resistance to midge. The first three principal components (PC) accounted for 56.22\% of the total genotypic variation. Based on the similarity of traits within and between members of clusters, genotypes were divided into 18 clusters. Genetic variability was present for the traits under study in the tested germplasm. Hybridization of midge-resistant genotypes from different clusters should be able to yield new genotypes combining high yield and other desirable yield-contributing traits.
\end{abstract}

\section{Introduction}

Sorghum (\textit{Sorghum bicolor} L. Moench) is a tropical cereal crop grown in a wide range of environments, where it is one of the staple foods of many people. It is the fifth most important cereal crop after wheat (\textit{Triticum aestivum}), maize (\textit{Zea mays}), rice (\textit{Oryza sativa}), and barley (\textit{Hordeum vulgare}). In the USA and Australia, sorghum is an important source of animal feed, either as grain or as forage (Reddy, Ashok Kumar, and Sanjana Reddy 2010). In Niger, sorghum ranks second in production area among cereal crops; pearl millet (\textit{Pennisetum glaucum}) being number one.
In Niger, sorghum production rarely meets the demand of the growing population. Grain sorghum yields are low. Not long ago, the average grain yield was about 0.280 ton/ha, which was far below the yield obtained in Argentina (4.9 tons/ha), the USA (4.3 tons/ha), and China (3.2 tons/ha) (FAOSTAT 2014). The low yield is mainly attributed to abiotic and biotic stresses. The major biotic stress is the sorghum midge (Stenodiplosis sorghicola), which is a panicle-feeding insect found in most of the sorghum-growing environments in Niger. Grain yield losses attributable to midge were 56% to 67% when local midge-susceptible sorghum varieties were grown in Niger compared with 14% to 17% yield losses when introduced, midge-resistant sorghum genotypes were grown (Kadi Kadi, Kapran, and Pendleton 2005).

To prevent sorghum midge infestations, farmers in Niger delay or advance sorghum planting date (Maman et al. 2004). Sorghum planting date must correspond with the onset of rains; however, it is difficult to predict the rainy season in the Sahelian countries, such as Niger. Jones and Johnson (1991) reported that effects of stress attributed to environmental factors on final yield depended on the growth stage in which stress occurred and the genotype planted. According to Vanderlip (1993), planting date significantly influenced grain yield in sorghum. Early or late planting tends to increase the risk of poor crop establishment, resulting in low yield with early planting or the occurrence of water stress or heat stress at reproductive stages with late planting (Vanderlip 1993).

The use of indigenous varieties with low-yielding capacity also limits sorghum productivity in Niger. Farmers mostly rely on low-yielding landraces, so sorghum production fails to meet demand of increasing population, and food insecurity remains a major challenge (Maman et al. 2004). Farmers need high-yielding and midge-resistant sorghum cultivars.

Suitable selection criteria for the identification of genotypes with desirable traits are essential for successful varietal improvement programs. Analysis of variability among traits and the identification of associations among various traits contributing to yield would facilitate successful development of high-yielding varieties (Mary and Gopalan 2006). However, selecting only for grain yield may not be efficient for developing varieties for adoption by farmers; selection, which integrates yield and farmer-preferred traits, should provide more appropriate varieties (Alvi et al. 2003). The identification of yield-related traits could result in more effective selection for yield and farmer-preferred traits.

Principal component analysis is a multivariate technique for examining the relationships among several quantitative variables (Johnson 2012). It is the most common technique used in variability studies and numerical classification; it is useful in grouping varieties based on their similarities (Bello 2004). Principal component analysis is an important breeding tool.
commonly used by breeders to identify traits that could be used to discriminate crop genotypes (Das 2000; Yan and Kang 2003). Establishing suitable selection criteria for identifying genotypes with desirable traits is useful in developing improved varieties. Analysis of variability among traits and knowledge of associations among traits contributing to yield would be of great importance in planning a successful breeding program (Mary and Gopalan 2006). To date, in Niger, no study has been carried out with the objective to assess diversity in sorghum based on traits mentioned above by using multivariate analysis. Therefore, the objective of this study is to determine genetic diversity of sorghum inbred lines, which would be helpful in enhancing the efficiency of sorghum breeding program.

Materials and methods

Experimental germplasm

Two-hundred and eighty F$_5$ recombinant inbred lines and two local checks were evaluated. The F$_5$ lines were obtained by crossing a local, white-grain sorghum variety (MDK), which is widely grown in Niger, with an exotic sorghum midge-resistant cultivar (ICSV88032) from International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). Progeny were advanced using single-seed descent (SSD). The local variety, MDK, possessed many farmer-preferred qualities; however, it lacked midge resistance and photo-insensitivity.

Experimental sites

The study was carried out during 2015 rainy season at two research stations of the Institut National de la Recherche Agronomique du Niger (INRAN) at Konni [latitude: 13° 47’23” North, longitude: 5° 14’ 57” East; average annual rainfall = 589.7 mm; and average temperature = 29.3°C] and Maradi [latitude: 13° 18’ 25” North, longitude: 7° 09’ 35” East; average annual rainfall = 537.4 mm; and average temperature = 20.5°C]. Both locations are considered hotspots for sorghum midge infestation in Niger.

Experimental design

The experiment was conducted using an alpha (0.1) lattice design with two replications at two different planting dates (PDs) at two different locations; PD-location combinations constituted four environments (E1 = first PD at Konni, E2 = second PD at Konni, E3 = first PD at Maradi, and E4 = second PD at Maradi). Blocking effects not being significant, data were re-analyzed using a randomized complete-block design. Each experimental unit consisted
of a single, 3-m long row, with intra-row spacing of 0.20 m and inter-row spacing of 0.80 m. The planted material was subjected to natural infestation by sorghum midge. To evaluate midge damage on the panicles, immediately after emergence, three panicles were covered with pollination bags. At harvest, panicle weight and grain weight were recorded for the three covered and three uncovered panicles. The loss in grain weight in the three uncovered panicles was expressed as percentage of grain weight of covered panicles.

Data collection and analysis

Data were collected on grain yield (GY), plant height (PH), 1000-seed weight (TSW), days to 50% flowering (FF), midge damage (MD), panicle compactness (PC), and awned glume (AG). Grain yield was measured in tons per hectare, adjusted to a grain moisture content of 12%. Plant height was measured in centimeters from the base of the plant to top of the panicle. Thousand-seed weight was measured as weight of 1000-seeds in grams. Days to 50% flowering was recorded by counting the number of days from planting to when 50% of the plants in a plot had flowered. Midge damage was calculated as loss of grain yield of three uncovered panicles expressed as percentage of grain yield of three covered panicles. This was calculated using the following formula adapted from (Sharma et al. 2003):

\[
MD = \frac{(X1 - X2)}{X1} \times 100
\]

where \(X1\) = grain yield of covered panicles and \(X2\) = grain yield of uncovered panicles.

Panicle compactness was recorded by visual observation of panicles at physiological maturity; a score of 1 represented compact panicles, 2 semi-compact panicles, and 3 non-compact panicles. Awned glume was scored as 1 and non-awned one as 2. Principal component analysis was conducted using XLSTAT to determine the relatedness of traits. Combined analysis of variance and hierarchical cluster analysis were performed in GenStat.

Results and discussion

Combined analysis of variance showed significant differences among genotypes for all traits under study (Table 1). This suggested that the germplasm was genetically diverse and considerable variation existed in the materials used. Such variation can be exploited by plant breeders in hybridization programs.

Three out of seven principal components with eigenvalues > 1 were extracted. These three components contributed 56.22% of the total variation among the germplasm. Principal components 1, 2, and 3 contributed,
respectively, 22.73%, 17.99%, and 15.50% toward the variation observed among genotypes (Table 2). The aim of principal component analysis is to resolve the total variation of a set of traits into linear, independent composite traits, which successively maximize variability in the data (Johnson 2012). Sample traits are generally inter-correlated to varying degrees and hence not all principal components are needed to summarize the data adequately. In this study, the first three principal components represented a sizeable amount of diversity among the genotypes investigated. This implied that several traits were involved in explaining the variation among the genotypes. Ayana and Bekele (1999) reported significance of first five PCs in the total variability of different agro-morphological traits in sorghum. The first four principal components, with eigenvalues greater than one, were also documented in 25 forage and 45 grain sorghum genotypes for dual purpose (Chikuta et al. 2015). Abraha et al. (2015) reported four principal components with eigenvalues greater than one, which explained > 75% of the total variation for grain yield, biomass, stay-green, leaf area, peduncle exertion, days to flowering, and maturity. Around 44%, 17%, and 15% variation

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>(^{†})PH</th>
<th>(^{‡})TWS</th>
<th>(^{§})FF</th>
<th>(^{¶})GY</th>
<th>(^{#})MD</th>
<th>(^{††})PC</th>
<th>(^{‡‡})AG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replications</td>
<td>1</td>
<td>191.20*</td>
<td>120.36ns</td>
<td>193.38ns</td>
<td>1,234,336*</td>
<td>0.735ns</td>
<td>0.034ns</td>
<td>0.026ns</td>
</tr>
<tr>
<td>Entries</td>
<td>281</td>
<td>9585.40**</td>
<td>102.64**</td>
<td>219.94**</td>
<td>177,377**</td>
<td>0.14**</td>
<td>0.08**</td>
<td>0.03**</td>
</tr>
<tr>
<td>Planting dates (PD)</td>
<td>1</td>
<td>1,287,836.10**</td>
<td>2005.55**</td>
<td>49.11ns</td>
<td>3,548,702**</td>
<td>50.42**</td>
<td>0.90**</td>
<td>3.32**</td>
</tr>
<tr>
<td>Sites</td>
<td>1</td>
<td>1,640,346.30**</td>
<td>100.86ns</td>
<td>12,805.54**</td>
<td>30,681,580**</td>
<td>27.80**</td>
<td>0.14ns</td>
<td>0.17**</td>
</tr>
<tr>
<td>Entries × PD</td>
<td>281</td>
<td>1471.00**</td>
<td>77.34ns</td>
<td>70.26**</td>
<td>129,768**</td>
<td>0.12**</td>
<td>0.04**</td>
<td>0.03**</td>
</tr>
<tr>
<td>Entries × Sites</td>
<td>281</td>
<td>1159.30**</td>
<td>80.28*</td>
<td>72.01**</td>
<td>119,234**</td>
<td>0.14**</td>
<td>0.04**</td>
<td>0.06**</td>
</tr>
<tr>
<td>PD × Sites</td>
<td>1</td>
<td>16,797.00**</td>
<td>3829.83**</td>
<td>379.18ns</td>
<td>7,723,780**</td>
<td>9.99**</td>
<td>3.67**</td>
<td>2.39**</td>
</tr>
<tr>
<td>Entries × PD × Sites</td>
<td>281</td>
<td>987.50ns</td>
<td>67.06ns</td>
<td>38.59ns</td>
<td>104,061**</td>
<td>0.11**</td>
<td>0.03**</td>
<td>0.02**</td>
</tr>
<tr>
<td>Residual</td>
<td>1127</td>
<td>852.4</td>
<td>64.47</td>
<td>49.91</td>
<td>62,286</td>
<td>0.055</td>
<td>0.025</td>
<td>0.008</td>
</tr>
</tbody>
</table>

* ** Significant at the 5% and 1% probability level, respectively. ns = non-significant.
† = Plant height (cm).
‡ = 1000-seed weight (g).
§ = Grain yield (t/ha).
¶ = Days to 50% flowering.
# = Midge damage (%).
†† = Panicle compactness (1 = compact panicles, 2 = semi-compact panicles and 3 = non-compact panicles).
‡‡ = Awned glume (1 = awned glume panicles and 2 = non-awned glume panicles).

Table 2. Principal component analysis of measured traits in 282 genotypes evaluated in 2015 rainy season.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
<th>PC7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>1.59</td>
<td>1.26</td>
<td>1.08</td>
<td>0.93</td>
<td>0.87</td>
<td>0.64</td>
<td>0.61</td>
</tr>
<tr>
<td>% Variance contribution</td>
<td>22.73</td>
<td>17.99</td>
<td>15.50</td>
<td>13.35</td>
<td>12.47</td>
<td>9.19</td>
<td>8.77</td>
</tr>
<tr>
<td>Cumulative variance contribution</td>
<td>22.73</td>
<td>40.72</td>
<td>56.22</td>
<td>69.57</td>
<td>82.04</td>
<td>91.23</td>
<td>100.00</td>
</tr>
</tbody>
</table>

respectively, 22.73%, 17.99%, and 15.50% toward the variation observed among genotypes (Table 2). The aim of principal component analysis is to resolve the total variation of a set of traits into linear, independent composite traits, which successively maximize variability in the data (Johnson 2012). Sample traits are generally inter-correlated to varying degrees and hence not all principal components are needed to summarize the data adequately. In this study, the first three principal components represented a sizeable amount of diversity among the genotypes investigated. This implied that several traits were involved in explaining the variation among the genotypes. Ayana and Bekele (1999) reported significance of first five PCs in the total variability of different agro-morphological traits in sorghum. The first four principal components, with eigenvalues greater than one, were also documented in 25 forage and 45 grain sorghum genotypes for dual purpose (Chikuta et al. 2015). Abraha et al. (2015) reported four principal components with eigenvalues greater than one, which explained > 75% of the total variation for grain yield, biomass, stay-green, leaf area, peduncle exertion, days to flowering, and maturity. Around 44%, 17%, and 15% variation
attributed to first, second, and third principal components, respectively, was reported by Chikuta et al. (2015). Several studies on principal component analysis of different agro-morphological traits in sorghum have been documented. Abraha et al. (2015) concluded that grain yield, biomass, stay-green, leaf area, peduncle exertion, days to flowering, and maturity were the most important traits for genetic variability in landrace sorghums. On the other hand, head width, head weight, grain yield per plant, and fresh and dry shoot weight were found to be the most important traits for drought tolerance in grain sorghum (Ali et al., 2011).

The phenotypic diversity observed in this study was attributable to several traits (Table 3). Variation relative to the first component was associated with plant height, 1000-seed weight, grain yield, and days to flowering. The second principal component was associated with days to flowering, panicle compactness, and midge damage. The third principal component was associated with awned glume and midge damage.

The 282 sorghum genotypes were grouped into 18 clusters based on similarity of morphological traits. The percentage of similarity was 64%, whereas the variances within and between clusters were 40.46% and 59.54%, respectively (Figure 1). The mean values of different traits in the different clusters ranged from 163 cm to 245 cm for plant height, 62 to 77 days after planting for days to 50% flowering, and 16.9 g to 28.3 g for 1000-seed weight. Grain yield of 6.46 t/ha was recorded and losses attributable to sorghum midge varied from 10% to 40%. In all the clusters, genotypes had semi-compact panicles, except in clusters 4 and 14, which had non-compact panicles and compact panicles, respectively. Most genotypes had awned glumes (Table 4). The cluster analysis substantiated that considerable variation existed among the genotypes. The significant genetic variability observed among the germplasm offers an excellent opportunity for genetic improvement of sorghum by hybridizing genotypes from different clusters to assemble desirable traits within varieties with high heterotic potential.

Attempts have been made to classify germplasm using morphological and physiological aspects in sorghum and other crops. In a study using

<table>
<thead>
<tr>
<th>Traits†</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PH</td>
<td>0.76</td>
<td>0.14</td>
<td>0.03</td>
</tr>
<tr>
<td>TSW</td>
<td>0.49</td>
<td>−0.49</td>
<td>−0.26</td>
</tr>
<tr>
<td>FF</td>
<td>0.44</td>
<td>0.64</td>
<td>0.23</td>
</tr>
<tr>
<td>GY</td>
<td>0.70</td>
<td>−0.25</td>
<td>0.28</td>
</tr>
<tr>
<td>MD</td>
<td>0.26</td>
<td>0.36</td>
<td>−0.59</td>
</tr>
<tr>
<td>PC</td>
<td>−0.11</td>
<td>0.61</td>
<td>0.21</td>
</tr>
<tr>
<td>AG</td>
<td>−0.02</td>
<td>−0.18</td>
<td>0.70</td>
</tr>
</tbody>
</table>

†PH = Plant height (cm); TSW = 1000-seed weight (g); FF = Days to 50% flowering; GY = Grain yield (t/ha); MD = Midge damage (%); PC = Panicle compactness (1 = compact panicles, 2 = semi-compact panicles and 3 = non-compact panicles); AG = Awned glume (1 = awned glume panicles and 2 = non-awned glume panicles).
microsatellites to characterize a collection of sorghum landraces, Burow et al. (2012) identified eight clusters based on genetic similarity of 0.5. They suggested that the results from the cluster analysis could help identify good parental materials for efficient use in breeding programs. The results of cluster analysis of 24 main sorghum traits of 72 sweet sorghum lines grouped the lines into four clusters (Feng et al. 2012). Similar results were found among 80 elite sorghum lines procured from ICRISAT (Jirali, Biradar, and...
Rao (2010). Ulaganathan and Nirmalakumari (2015) grouped 305 finger millet genotypes into 16 clusters and concluded that the clustering pattern could be utilized in choosing diverse genotypes that were likely to generate the highest possible variability for various economic traits. Similar results were reported by Mustafa et al. (2015) in maize and by Beheshtizadeh et al. (2013) in bread wheat.

**Conclusions**

We concluded that significant diversity existed among sorghum lines for the traits studied. Efficient exploitation of this diversity is helpful in identifying sorghum parental lines for hybrid breeding program. Inter-crossing between genotypes from diverse clusters will provide segregating sorghum progeny for yield and farmer-preferred traits; thereby, leading to the development of high-yielding varieties preferred by farmers.

**Acknowledgments**

The authors acknowledge Alliance for Green Revolution in Africa (AGRA) for funding this work. This work has also been undertaken as part of the CGIAR Research Program on Dryland Cereals.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This work was supported by the Alliance for Green Revolution in Africa (AGRA) and the CGIAR Research Program on Dryland Cereals.

**References**


