2006

Variation in the U.S. Photoperiod Insensitive Sorghum Collection for Chemical and Nutritional Traits

Tisha L. Hooks
University of Nebraska - Lincoln

Jeffrey F. Pedersen
University of Nebraska - Lincoln, jpedersen1@unl.edu

D. B. Marx
University of Nebraska-Lincoln, david.marx@unl.edu

K. P. Vogel
University of Nebraska-Lincoln, kvogel1@unl.edu

Published online February 24, 2006

Variation in the U.S. Photoperiod Insensitive Sorghum Collection for Chemical and Nutritional Traits

Tisha Hooks, J. F. Pedersen,* D. B. Marx, and K. P. Vogel

ABSTRACT

Screening germplasm for chemical and nutritional content can be expensive and time consuming. Near infrared spectroscopy (NIRS) and application of geostatistical models can make screening more efficient. The objectives of this study were to utilize these technologies to: (i) generate chemical and nutritional values for the U.S. photoperiod insensitive sorghum collection, (ii) describe variability for those traits, (iii) identify accessions in the highest and lowest 1% for each trait, and (iv) describe relationships among the accessions. Accessions were grown at Ithaca, NE, during 2001 and 2002. Samples of grain were scanned and NIRS equations developed for starch, fat, crude protein, acid detergent fiber, and phosphorus. The NIRS generated values for each accession can be accessed on GRIN at http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?69; verified 22 November 2005. The highest and lowest 1% of accessions was identified for each trait by best linear unbiased predictors (BLUPs). Means and standard deviations for observed values and variances due to accessions were calculated. Rank correlations between BLUPs and observed values ranged from $r = 0.82$ to $r = 0.92$. Principal component analysis showed that much of the variation is attributable to a contrast of starch with a weighted average of fat, crude protein, acid detergent fiber, and phosphorus. Cluster analyses showed clusters on the basis of canonical values, but no geographical, taxonomical, or morphological interpretation of the clusters was apparent.

SCREENING large germplasm collections for quantitative traits such as chemical and nutritional content can be prohibitively expensive and time consuming by traditional experimental designs and analytical techniques. For example, the current U.S. sorghum collection contains over 42 000 accessions (USDA-ARS, 2005a). Growing each sorghum accession in one single-row plot 7.6 m long spaced 76 cm apart would require over 24 ha of research plots per replication. Total per sample analysis costs for crude protein (CP), acid detergent fiber (ADF), phosphorus (P), calcium (Ca), starch, fat, and various calculated values can total $35/sample for wet chemistry (Ward Laboratories, 2005). Few research projects have the resources to undertake screening of such a large collection by traditional replicated field studies and wet chemical analysis.

Two technologies, NIRS and the use of new statistical models made possible by the increasing speed and power of computers, can be used to generate predicted values for chemical and nutritional content on the basis of a relatively small subset of samples and to account for spatial and temporal variability, reducing the need for replication of accessions or replicated check plots in augmented designs. NIRS predicts chemical and nutritional values from spectral data and is used to generate chemical and nutritional values for grain and forage samples. Time required to scan a sample is approximately 1 min, and cost per sample is low compared with the wet chemistry methods. The technology is commonplace in many analytical labs and has been approved for use in testing ADF and protein in feeds for over a decade (AOAC, 1990). More recently, it was approved for testing of corn grain (*Zea mays* L.) for oil, protein, and starch content by the USDA Grain Inspection, Packers and Stockyards Administration (USDA-GIPSA, 1999).

Numerous statistical approaches to make large early generation or germplasm screening efforts more efficient have been developed during the past century. Augmented designs (Federer, 1956; Steel, 1958; Searle, 1965) and nearest neighbor analysis (Papadakis, 1937; Bartlett, 1937) are both examples that have been widely used. However, neither offered the needed increase in efficiency for large germplasm screening experiments, and neither specified the nature of the relationship between neighboring plots. Applications of geostatistical models to account for various correlation structures in field experiments have led to increases in accuracy and precision (Brownie et al., 1993; Cullis and Gleeson, 1991; Zimmerman and Harville, 1991). Mixed model equations developed by Henderson (1975) have become useful with the widespread availability of analytical software (Littell et al., 1996) and more powerful computers, and they have proven useful for the analysis of spatially correlated data (Henderson, 1975; Marx and Stroup, 1993).

Recent research in our laboratory using PROC MIXED (SAS, 2003) and simulated data to compare the efficiency of germplasm screening experiments with varying levels of check plots demonstrated that the use of BLUPs was superior to the use of least square means (LS means) and observed values for selecting the highest proportion of true top ranking genotypes and that BLUPs were influenced little by the proportion of check plots to experimental plots (Sebolai et al., 2005). Incorporation of the correlation structure and the use of BLUPs to

Tisha Hooks and D.B. Marx, Dep. of Statistics, Univ. of Nebraska-Lincoln, Lincoln, NE 68583; J.F. Pedersen and K.P. Vogel, USDA-ARS, NPA Wheat, Sorghum and Forage Research, 344 Keim Hall, Univ. of Nebraska-Lincoln, Lincoln, NE 68583-0937.Joint contribution of the USDA-ARS and the Univ. of Nebraska Agric. Exp. Stn. As Paper no. 14532, Journal Series, Nebraska Agric. Exp. Stn. Received 4 May 2005. *Corresponding author (jfp@unlserve.unl.edu).

Published in Crop Sci. 46:751–757 (2006).
Plant Genetic Resources
doi:10.2135/cropsci2005.05-0047

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: ADF, acid detergent fiber; BLUPs, best linear unbiased predictors; Ca, calcium; CP, crude protein; GRIN, Germplasm Resources Information Network; NIRS, near infrared spectroscopy; P, phosphorus; SEC, standard error calibration; SECV, standard error cross validation.

select superior accessions may be especially well suited to screening germplasm collections since they were originally developed for ranking and selection (Robinson, 1991). The use of BLUPs is, however, premised on treatment (accession) effects being assumed random. In screening experiments involving a subpopulation of accessions from a collection, this would be valid if the subsetting criteria were independent of screening criteria.

A subset of the U.S. sorghum collection that are of particular interest for sorghum improvement is comprised of approximately 4000 photoperiod insensitive accessions that will flower at temperate latitudes. The objectives of this study were to (i) generate chemical and nutritional values for grain from the U.S. photoperiod insensitive sorghum collection, (ii) describe the variability for those chemical and nutritional values, (iii) identify accessions in the highest and lowest 1% for each trait after accounting for spatial and temporal variation, and (iv) describe relationships among the accessions using these values.

MATERIALS AND METHODS

Approximately 75% of the accessions in the U.S. photoperiod insensitive sorghum collection were grown during 2001 and 2002 at the University of Nebraska Field Laboratory, Ithaca, NE, on a predominantly Sharpsburg silty clay loam (fine, montmorillonitic, mesic Typic Arguidolls) site. Plots consisted of a single 7.6-m row of each accession spaced 76 cm apart, and individual accessions were planted in only 1 yr. The variety 'Wheatland' was interspaced throughout the nurseries at a density of 16% of the plots. Each plot was seeded with a precision vacuum planter calibrated to deliver 120 seeds per row (240 000 seeds ha^{-1}). Because of limitations on field space, 1990 accessions were planted in 2001 and 1215 accessions were planted in 2002. Germination was generally adequate for plot establishment. Plots that did not germinate and those that did not reach maturity were treated as missing plots. Following the 2002 harvest, the decision was made that the 2914 accessions already grown in this experiment should adequately represent the diversity of the U.S. photoperiod insensitive sorghum collection and the remaining lines were not planted in subsequent years. Accessions grown in 2001 or 2002 were not preselected for any traits or descriptors and therefore can be considered a random subset of the photoperiod insensitive sorghum collection.

The experiments were planted 19 May 2001 and 23 May 2002. Nitrogen fertilizer was applied preplant at both locations at 157 kg ha^{-1}. Atrazine (6-chloro-*N*^{2}-ethyl-*N*^{4}-isopropyl-1,3,5-triazine-2,4-diamine) was applied at 2.2 kg ha^{-1} immediately after planting, followed by an application of quinclorac (3,7-dichloro-8-quinolinecarboxylic acid) and atrazine at 0.37 kg ha^{-1} and 1.1 kg ha^{-1}, respectively, approximately 14 d post emergence for weed control. In 2002, bentazon [3-(1-methylethyl)-1*H*-2,1,3-benzothiadiazin 4(3*H*)-one-2,2-dioxide] was added to the post emergence application at 0.28 kg ha^{-1} for velvetleaf [*Abutilon theophrasti* (Medik)] control. No supplemental irrigation was applied in 2001. Plots were irrigated by furrow irrigation on 2 July and 23 July in 2002.

Ten open pollinated panicles per plot were harvested as accessions reached maturity, bulked by row, threshed, and seed were stored at 10°C for NIRS scanning and chemical and nutritional quality analysis. All samples were ground to pass a 1-mm screen on a cyclone mill and scanned on a Model 6500 near infrared reflectance spectrometer (NIRS Systems, Silver Spring, MD).[1] To cover the wide expected range in spectral diversity, 541 reference samples (50% from each year) were selected by cluster analysis of the reflectance data for wet chemistry analysis (Shenk and Westerhaus, 1991). Standard wet chemistry methods were used to determine starch (Y S I, 2000), fat (Padmore, 1990a), CP (Miller et al., 1998), ADF (ANKOM, 1999), calcium (Padmore, 1990b), and phosphorus (Padmore, 1990b) content of reference samples. All lab analyses were performed by Ward Laboratories, Kearney, NE. The wet chemistry values were then used to develop NIRS prediction equations by partial least squares (Shenk and Westerhaus, 1991) and to generate observed values for each row in each year. The calibration statistics for each trait are shown in Table 1.

Table 1. Calibration statistics for NIRS† prediction of starch, fat, CP, ADF, calcium, and phosphorus.

Trait	Math treatment‡	R^2	SEC	SECV
			g kg^{-1}	
Starch	2,6,4,1	0.69	0.153	0.161
Fat	3,5,5,1	0.75	0.045	0.047
CP	1,4,4,1	0.85	0.066	0.070
ADF	1,4,4,1	0.75	0.081	0.075
Ca	1,4,4,1	0.35	0.001	0.001
P	1,4,4,1	0.73	0.003	0.004
TDN	1,4,4,1	0.74	0.126	0.133

† Abbreviations: NIRS = near-infrared spectroscopy, CP = crude protein, ADF = acid detergent fiber, Ca = calcium, *P* = phosphorus, SEC = standard error calibration, SECV = standard error cross validation.

‡ Math treatments used by NIRS software = derivative number, gap (nm over which derivative is calculated), smooth (number of points over which data is smoothed), second smooth (number of points over which data is smoothed a second time).

The mean and standard deviation of the observed values for each trait were calculated for each year separately and also for the combined data set. To identify the underlying spatial variation for each response variable, SAS PROC VARIOGRAM was used to calculate a semivariogram for each of the chemical and nutritional traits (assuming a no-nugget model) in both 2001 and 2002. Since the default initial values in SAS PROC MIXED often lead to unreasonable estimates for the range and sill (Marx and Stroup, 1993), the range and sill of the semivariograms were estimated visually to obtain feasible estimates for use as starting values. Then, analyses incorporating various known correlation structures were run for each year in PROC MIXED, and the model fitting information was used to determine which correlation structure best fit the data for each year separately (Littell et al., 1996). It was determined that an exponential model provided the best fit in all cases. Also, since it was anticipated that the variability would differ between the 2 yr, the geostatistical model with accessions regarded as random effects includes a separate accession variance, range, and sill for each year. The model is as follows:

$$y_{ijk} = \mu + Y_k + \tau_{i(k)} + e_{ijk}, \text{ with } \tau_{i(k)} \sim iid\ N(0, \sigma^2_{\tau_k}) \text{ and}$$

$$e_{ijk} \sim iid\ N\left[0, \sigma^2_{e_k} \exp\left(\frac{-d_{ij}}{\rho_k}\right)\right],$$

where y_{ijk} represents the chemical trait of interest, μ is the overall intercept, Y_k is the effect of the kth year, $\tau_{i(k)}$ is the effect of the ith accession in the kth year, e_{ijk} is the random

[1] The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the United States Department of Agriculture, the Agricultural Research Service, or the Northern Plains Area of any product or service to the exclusion of others that may be suitable.

Table 2. Mean, standard deviation, and range of observed NIRS chemical and nutritional values of sorghum photoperiod insensitive accessions and Wheatland.

	ADF†	CP	Fat	P	Starch
2001 Sorghum accession statistics					
Mean	**0.60**	**1.30**	**0.44**	**0.04**	**6.92**
StDev	**0.14**	**0.15**	**0.08**	**0.01**	**0.13**
Min	**0.30**	**0.92**	**0.25**	**0.02**	**6.41**
Max	**1.35**	**1.86**	**0.71**	**0.07**	**7.25**
2002 Sorghum accession statistics					
Mean	**0.61**	**1.27**	**0.39**	**0.04**	**7.05**
StDev	**0.12**	**0.14**	**0.06**	**0.01**	**0.12**
Min	**0.28**	**0.89**	**0.24**	**0.03**	**6.27**
Max	**1.65**	**2.13**	**0.64**	**0.06**	**7.37**
2001 Wheatland statistics					
Mean	**0.61**	**1.29**	**0.45**	**0.04**	**6.91**
StDev	**0.06**	**0.15**	**0.10**	**0.01**	**0.11**
2001 Wheatland statistics					
Mean	**0.52**	**1.18**	**0.36**	**0.04**	**7.15**
StDev	**0.05**	**0.09**	**0.02**	**0.01**	**0.06**

† Abbreviations: CP = crude protein, ADF = acid detergent fiber, Ca = calcium, *P* = phosphorus, and StDev = standard deviation.

error term, d_{ij} is the distance between observation i and observation j, $\sigma^2_{\epsilon_k}$ is the sill of the semivariogram for Year k, and ρ_k is the range of the semivariogram for Year k.

This model has two variance components for accession, $\sigma^2_{\tau_1}$ and $\sigma^2_{\tau_2}$, which are variances of the accessions within Year 1 (2001) and Year 2 (2002), respectively. These values are a measure of the variability that exists across all possible accessions for each chemical trait in each year.

The entire data set was analyzed by incorporating the above statistical model into SAS PROC MIXED. Again, since accessions were regarded as random effects, the MIXED procedure calculates BLUPs which are adjusted for underlying spatial variability. The BLUPs were output into a separate data set, and SAS PROC RANK was used to obtain the highest and lowest 1% of the accessions on the basis of the values of the BLUPs. SAS PROC RANK was also used to obtain the highest and lowest 1% of the accessions on the basis of the observed values of the chemical and nutritional traits in order to determine if the results changed significantly after accounting for spatial variability. Since primary interest lies in the ranking of the accessions, Spearman's rank correlation between the BLUPs and the observed values was calculated for each chemical trait.

In addition to describing the variability present in this sorghum collection, it was of interest to describe relationships that exist among the accessions and also between the various chemical traits. Principal components for this particular analysis were defined as follows: **S** is the sample variance-covariance matrix of the following random variables: starch, fat, CP, ADF, and P. The first principal component is a linear combination $\boldsymbol{a}_1'\boldsymbol{x} = a_{11}x_1 + \cdots + a_{15}x_5$ that maximizes the variance among all linear combinations of **x** subject to $a_1'a_1 = 1$. The second principal component $\boldsymbol{a}_2'\boldsymbol{x}$ with $a_2'a_2 = 1$ is such that it is uncorrelated with the first and its variance is highest among all the linear combinations uncorrelated with the first principal component. Similarly, the third, fourth, and fifth principal components, all uncorrelated with the others, are defined. Then $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_5 > 0$ are the eigenvalues and $\mathbf{a}_1$, …, $\mathbf{a}_5$ become the corresponding eigenvectors of **S**, where $\boldsymbol{a}_i'\boldsymbol{a}_i = 1$ for i = 1,2, … 5. The values $\boldsymbol{a}_1'\boldsymbol{x}$, $\boldsymbol{a}_2'\boldsymbol{x}$, …, $\boldsymbol{a}_5'\boldsymbol{x}$ are precisely the first, second, …, fifth principal components of **x**, and the eigenvalues of **S** are the variances of the corresponding principal components.

Table 3. Sorghum photoperiod insensitive accessions identified with high and low acid detergent fiber (ADF) using BLUPS and observed NIRS values.

Based on BLUPS		Based on NIRS values	
Highest 1%	**Lowest 1%**	**Highest 1%**	**Lowest 1%**
Plant introduction accession number			
591372†	**563622**	**591372**	**563731**
550635	**563720**	**535800**	**563614**
535776	**174377**	**177158**	**Grif552**
535800	**563680**	**177078**	**Grif530**
302120	**563623**	**302120**	**550851**
246593	**563590**	**550635**	**563601**
177158	**170785**	**246593**	**542741**
550616	**560396**	**535799**	**563769**
179051	**177077**	**174381**	**563594**
177078	**170792**	**535776**	**563707**
285002	**563743**	**550616**	**563721**
302123	**563778**	**302123**	**563720**
174381	**563616**	**179051**	**Grif600**
602677	**563625**	**535802**	**563600**
602696	**563615**	**173112**	**563622**
601916	**563769**	**173113**	**563674**
601911	**563733**	**550627**	**563685**
173116	**563601**	**177158**	**550680**
602678	**563685**	**285002**	**563615**
204631	**563721**	**535803**	**Grif582**
599750	**563837**	**170800**	**550815**
535778	**563635**	**173116**	**563635**
601907	**563707**	**535798**	**563625**
550627	**563594**	**170802**	**563630**
601898	**563600**	**535801**	**563778**
535779	**563731**	**535778**	**550864**
601894	**174378**	**535779**	**Grif502**
314743	**563614**	**535772**	**Grif521**
Accession Variance2001 = 0.0095 (g kg^{-1})^2 2002 = 0.0132 (g kg^{-1})^2			
Spearman's Correlation of all BLUPS with NIRSr = 0.91			

† Accessions with numbers between 560376 and 602840 were grown in 2002. All other accessions were grown in 2001.

Table 4. Sorghum photoperiod insensitive accessions identified with high and low crude protein (CP) using BLUPS and observed NIRS values.

Based on BLUPS		Based on NIRS values	
Highest 1%	**Lowest 1%**	**Highest 1%**	**Lowest 1%**
Plant introduction accession number			
591372†	**602767**	**591372**	**601841**
601827	**602806**	**513645**	**308484**
562744	**563811**	**465236**	**601877**
602669	**601941**	**248270**	**542733**
602711	**563627**	**248272**	**563731**
513645	**563565**	**206994**	**267385**
567930	**601935**	**291211**	**601876**
602721	**563751**	**511790**	**601935**
585317	**563733**	**527429**	**602750**
562769	**599831**	**527428**	**526638**
602839	**586385**	**205230**	**563614**
602743	**562748**	**302123**	**542736**
567914	**563819**	**562744**	**601931**
602668	**601930**	**223719**	**601936**
602831	**563212**	**532882**	**Grif530**
563698	**563625**	**291382**	**550682**
602667	**586453**	**206750**	**601930**
599930	**563635**	**179502**	**563593**
586463	**602749**	**164416**	**563625**
560409	**573273**	**586463**	**602748**
567807	**601841**	**527427**	**563212**
599932	**563837**	**291381**	**586385**
602833	**562714**	**302134**	**563707**
602713	**602750**	**71309**	**563794**
586532	**563707**	**Grif622**	**563811**
602797	**563731**	**248269**	**565172**
599717	**565172**	**602711**	**586453**
599928	**563614**	**314743**	**527151**
Accession Variance2001 = 0.0081 (g kg^{-1})^{2}2002 = 0.0162 (g kg^{-1})^2			
Spearman's Correlation of all BLUPS with NIRSr = 0.92			

† Accessions with numbers between 560376 and 602840 were grown in 2002. All other accessions were grown in 2001.

Table 5. Sorghum photoperiod insensitive accessions identified with high and low fat using BLUPS and observed NIRS values.

Based on BLUPS		Based on NIRS values	
Highest 1%	Lowest 1%	Highest 1%	Lowest 1%
Plant introduction accession number			
567803†	602673	547917	576127
567930	599843	548038	Grif521
567811	550603	Grif7261	540511
567931	601838	547994	599878
586463	565172	547918	563417
567809	565128	567803	565142
567938	545586	542755	599837
586532	601865	567811	599739
568016	599739	542724	599839
567915	563277	547967	601865
563532	563657	90267	602679
602824	563448	547912	563453
567918	562174	547931	599936
567954	576126	567930	599876
568011	602679	547997	601922
586529	563416	548032	563657
563845	568053	548037	563448
567951	569812	302257	563416
586526	563068	547921	599843
567985	585370	548026	599970
563550	599878	547914	576126
567945	565142	542760	601822
567807	562777	548016	601853
563580	563453	547919	Grif530
563595	540511	547950	563449
563830	543253	547998	601821
586431	563417	567931	599836
567919	576127	547935	585370

Accession Variance2001 = 0.0019 (g kg^{-1})^2 2002 = 0.0033 (g kg^{-1})^2
Spearman's Correlation of all BLUPS with NIRSr = 0.87

† Accessions with numbers between 560376 and 602840 were grown in 2002. All other accessions were grown in 2001.

Table 6. Sorghum photoperiod insensitive accessions identified with high and low phosphorus (P) using BLUPS and observed NIRS values.

Based on BLUPS		Based on NIRS values	
Highest 1%	Lowest 1%	Highest 1%	Lowest 1%
Plant introduction accession number			
601827†	599842	513645	302120
599717	563733	550630	177078
563578	563837	527429	177158
602743	602806	550631	601935
602668	563707	550877	599707
602721	599860	164416	601835
602741	177158	302134	601841
567930	602750	455141	562714
599735	602686	475058	601936
563567	586071	527428	601866
562946	563819	532882	601865
602669	565163	533954	174379
563561	563614	550615	204919
601828	563721	206994	246593
562744	563764	291379	599861
599716	601841	302263	599860
567807	565159	465236	599843
560409	599879	586463	601832
586463	599861	71309	601840
567914	601935	291377	601837
602713	563731	302257	Grif552
585316	563565	533990	196891
601820	599707	542755	267385
599718	302120	543253	287573
563698	567971	545546	563731
586532	563627	548038	602686
602711	562714	548046	599879
601907	565172	549179	599842

Accession Variance2001 = 0.0000097 (g kg^{-1})^{2}2002 = 0.000018 (g kg^{-1})^2
Spearman's Correlation of all BLUPS with NIRSr = 0.91

† Accessions with numbers between 560376 and 602840 were grown in 2002. All other accessions were grown in 2001.

The first step in determining the principal components of a vector **x** was to obtain the eigenvalues and the eigenvectors of the sample variance-covariance matrix **S** of **x**. It is equally valid to start with a sample correlation matrix **R** instead of a covariance matrix, and this is actually recommended if measurements on different variables are on different scales and the variances are of different magnitudes (Johnson and Wichern, 1998). For this reason, the principal components in this analysis were obtained from the sample correlation matrix using SAS PROC PRINCOMP.

Finally, cluster analysis was conducted with descriptor data obtained from GRIN (USDA-ARS, 2005b) to search for "natural" groupings. Descriptor data included country of origin, endosperm color, endosperm texture, endosperm type, kernel color, kernel plumpness, kernel shape, panicle shape, pericarp color, and race. Country of origin data was limited to seven countries in Africa (Botswana, Ethiopia, Niger, Nigeria, South Africa, Sudan, and Zimbabwe) with large numbers of accessions in the U.S. photoperiod insensitive collection. The accessions were clustered on the basis of their standardized BLUP values for starch, fat, CP, ADF, and P. The clusters were formed using the K-means method implemented in SAS PROC FASTCLUS. This is a nonhierarchical clustering technique designed to group items into a collection of K clusters, and this technique is particularly useful for large data sets (Johnson and Wichern, 1998).

RESULTS

The NIRS chemical and nutritional values for each accession grown in this study are available on GRIN

Table 7. Sorghum photoperiod insensitive accessions identified with high and low starch using BLUPS and observed NIRS values.

Based on BLUPS		Based on NIRS values	
Highest 1%	Lowest 1%	Highest 1%	Lowest 1%
Plant introduction accession number			
174378†	550627	562714	591372
563614	535775	568031	535800
565172	535798	563731	535799
562714	527429	563614	302123
302135	179502	562748	550635
562748	302257	586453	535802
563249	181898	563249	177078
563731	177158	563212	550616
246594	205230	308484	177158
568031	567930	601936	535776
586453	527428	567971	535798
164903	535799	563269	246593
181084	285002	563593	170802
475589	271619	565172	170788
527151	179051	563751	535803
170785	535778	563625	179502
287693	535779	563648	170800
567971	204631	586076	205230
563212	173121	585335	550627
170792	602668	563210	179051
565164	550616	563579	173113
573273	170788	573273	174381
586560	246593	563395	535801
563764	535776	563819	527428
563819	550635	586560	535778
550603	302123	563752	179501
534152	535800	601876	173121
563635	591372	563685	535779

Accession Variance2001 = 0.023 (g kg^{-1})^{2}2002 = 0.022 (g kg^{-1})^2
Spearman's Correlation of all BLUPS with NIRSr = 0.82

† Accessions with numbers between 560376 and 602840 were grown in 2002. All other accessions were grown in 2001.

Table 8. Coefficients for principal components.

Variable	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$
Starch	−.536563	0.124615	0.190480	0.451646	0.675503
Fat	0.280836	0.734362	−.580312	0.084303	0.194872
CP†	0.519711	0.007863	0.397989	−.455223	0.603503
ADF	0.361604	−.641427	−.483980	0.369597	0.294915
P	0.482053	0.183557	0.484070	0.667142	−.233516
Variance ($\hat{\lambda}_i$)	2.96639	1.01009	0.638674	0.271437	0.113407
Cumulative percentage of total variance	59.33	79.53	92.30	97.73	100.00

† Abbreviations: CP = crude protein, ADF = acid detergent fiber, *P* = phosphorus.

(USDA-ARS, 2005b) and can be accessed at the following URL: http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?69' verified 22 November 2005. Prediction equations for Ca were judged to have too low of an R^2 to be useful and Ca values are not reported. The mean, standard deviation, and range of NIRS values of accessions and Wheatland are shown in Table 2. In 2001, mean accession values were similar to Wheatland. In 2002, mean accession values for ADF, CP, and Fat appeared slightly higher, and the mean accession values for other traits were similar or slightly lower than the Wheatland mean. The maximum range of values for most traits was greater in 2002, possibly due to irrigation allowing the accessions to express more of their genetic potential.

The highest and lowest 1% of accessions identified for each trait using BLUPs and using the observed NIRS values is shown in Tables 3–7. Variance associated with accessions after accounting for various spatial structures in the fields are shown for each year. Spearman rank correlation coefficients for BLUPS with NIRS values are also shown. The rank correlation coefficients ranged from 0.91 to 0.92 for ADF, CP, and P, and considerable redundancy existed between the accessions selected using BLUPs and observed NIRS values. The rank correlations for fat, and starch were 0.87 and 0.82, respectively.

The principal component analysis is summarized in Table 8. The first and second principal components collectively account for 79.5% of the total variability. Consequently, the sample variation is adequately summarized by only two principal components. These principal components are given as follows:

$$PC_1 = -0.536563 \times \text{starch} + 0.280836 \times \text{fat} + 0.519711 \times \text{cp} + 0.361604 \times \text{ADF} + 0.482053 \times \text{phosphorus}$$

$$PC_2 = 0.124615 \times \text{starch} + 0.734362 \times \text{fat} + 0.007863 \times \text{cp-}0.641427 \times \text{ADF} + 0.183557 \times \text{phosphorus}.$$

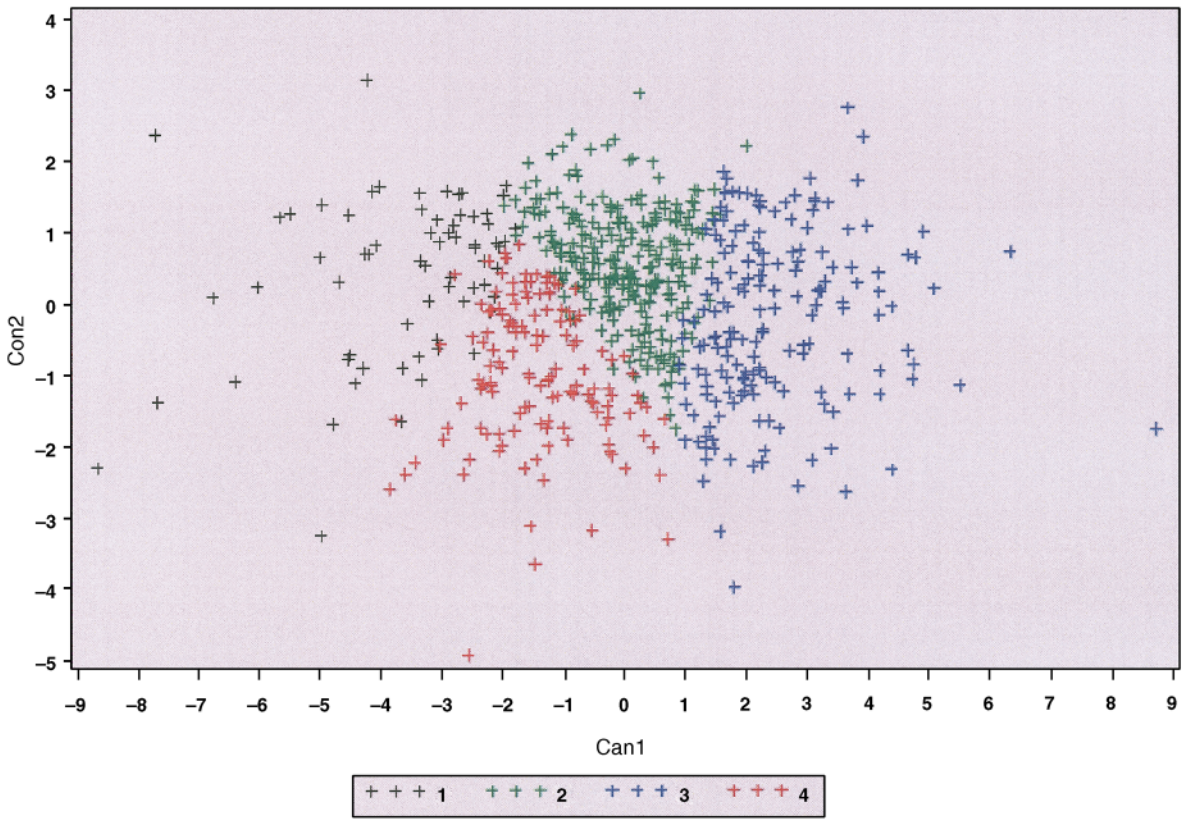

Fig. 1. Clusters based on BLUPs for selected countries in Africa.

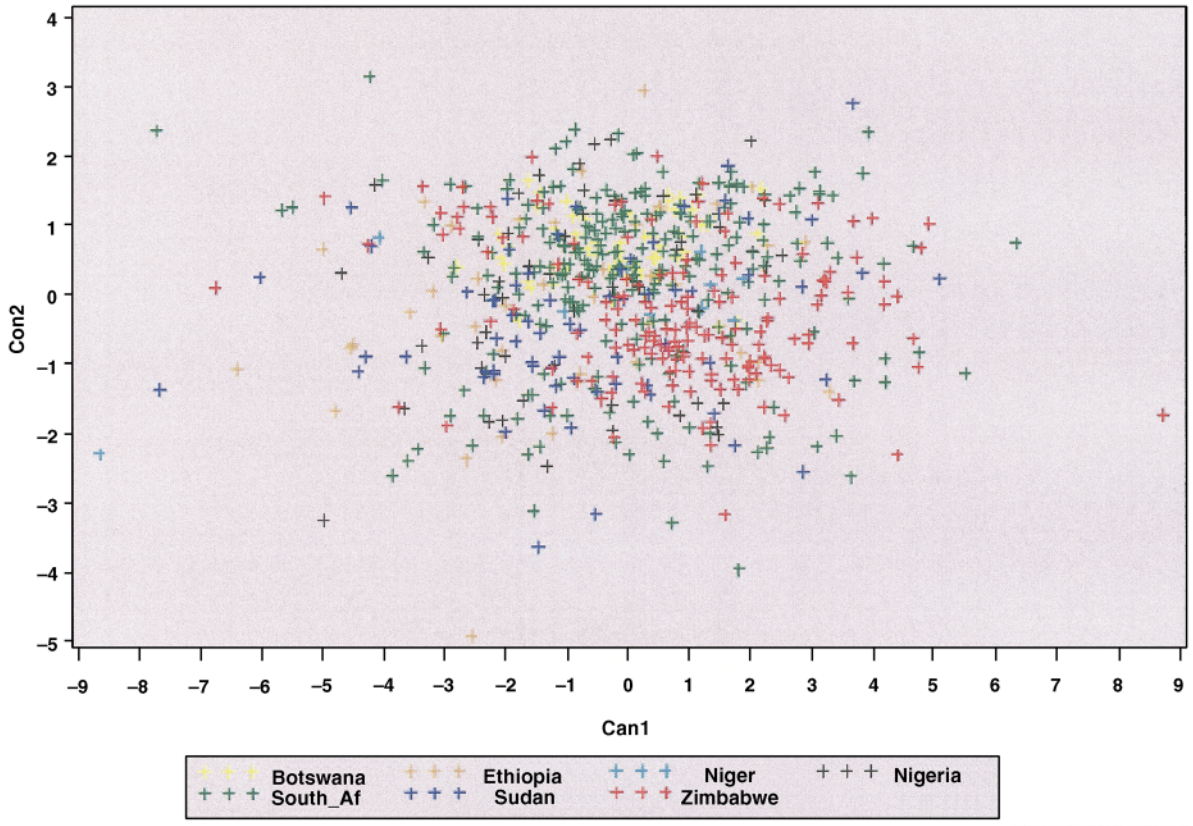

Fig. 2. Data plotted by country of origin.

Four clusters were formed in the analysis. SAS PROC CANDISC was used to compute the canonical variables for plotting the clusters, and the results are shown (Fig. 1). Canonical variables were also plotted on the descriptors obtained from GRIN (USDA-ARS, 2005b) for each point. No clustering was attributable to country of origin endosperm color, endosperm texture, endosperm type, kernel color, kernel plumpness, kernel shape, panicle shape, pericarp color, or race. A sample plot of canonical variables and country of origin are shown (Fig. 2).

DISCUSSION

As apparent from the lists of selected accessions, the use of BLUPs causes differences in accessions identified in the highest or lowest 1% for a given trait when compared with the use of observed values. BLUPs account for spatial variation; thus, their rankings should differ from the rankings of the observed values. However, drastic differences should not exist between the two sets of rankings, resulting in high correlation coefficients between BLUPs and observed values. This is indeed what we observed in the sorghum photoperiod insensitive collection, with rank correlation coefficients ranging from $r = 0.82$ to 0.92.

The importance of these difference and similarities was demonstrated in simulations by Sebolai et al. (2005) who concluded that in relatively uniform fields, the use of BLUPs identified 56 to 60% of the true top ranking accessions, while the use of the observed values identified only 35 to 37% of the true top ranking accessions in a population and that both were relatively unaffected by check plot density. While it is not possible within the constraints of this field screening experiment to determine if more of the true highest and true lowest 1% of the accessions were selected for any trait using BLUPs or observed values, the results from Sebolai et al. (2005) strongly suggest that basing selection on BLUPs will at least equal selection using observed values in nonuniform field conditions and double the probability of identifying accessions with true high and low values for a given trait under uniform field conditions such as those common to research fields. We therefore recommend the use of the selections based on BLUPs (Tables 3–7) to identify probable "best" and "worst" accessions for chemical and nutritional traits within this population of accessions for future use by breeders.

Principal component analysis often reveals relationships that may not have been previously suspected and thus allows interpretation that would not normally result (Johnson and Wichern, 1998). In many cases, it is possible to give meaningful interpretations to the coefficients (eigenvectors) in the principal components. The analysis indicated that nearly 80% of the sample variation is described by the first two principal components. On the basis of its coefficients, the first principal component appears to contrast starch content with a weighted average of fat, crude protein, acid detergent fiber, and

phosphorus content. The second principal component appears to be a contrast between fat and acid detergent fiber content (since the coefficients for the other variables are much smaller in magnitude, these variables can be ignored in the definition of the second principal component). Since starch content represents the largest chemical fraction of sorghum grain, its importance in contributing to sample variation should come as no surprise.

An alternative approach to understanding the results of the principal component analysis is to consider them on a morphological basis. Starch is found primarily in endosperm. Fat and protein are found primarily in the embryo (or "germ") fraction of the seed. Fiber is found primarily in the seed coat. One might therefore infer that the first principal component is essentially a contrast of endosperm vs. all other seed fractions and that the second component is a contrast of germ vs. seed coat. In both cases, variation is explained by contrasting the energy-rich components of sorghum grain with less energy-rich components. Understanding energy storage and compartmentalization, therefore appears key to understanding variation among accessions in this collection.

The plot of canonical values shown in Fig. 1 clearly displays clustering of the values. These clusters suggest relationships among the accessions. However, when canonical values are plotted by country of origin (Fig. 2), endosperm color, endosperm texture, endosperm type, kernel color, kernel plumpness, kernel shape, panicle shape, pericarp color, and race, no geographical, morphological, or taxonomic interpretation can be assigned to the clusters identified in Fig. 1.

In conclusion, chemical and nutritional traits in the U.S. photoperiod insensitive sorghum collection were characterized by NIRS and statistical techniques to account for the effects of spatial variability common to field screening nurseries. These combined technologies allowed us to identify the highest and lowest 1% of the accessions for five chemical and nutritional traits. Principal component analyses attributed much of the total variation among the accessions to a contrast of starch with a weighted average of fat, crude protein, acid detergent fiber, and phosphorus. Cluster analyses showed clear separation of clusters on the basis of canonical values, but no geographical, taxonomical, or morphological interpretation of the clusters was apparent.

ACKNOWLEDGMENTS

The authors thank Mr. Steven Masterson for his efforts in developing NIRS prediction equations and generating the NIRS data used in this study.

REFERENCES

A.N.K.O.M. Technology. 1999. Method for determining acid detergent fiber, Ankom 200/220 Fiber Analyzer. Ankom Technology, Fairport, NY.

AOAC. 1990. Fiber (acid detergent) and protein (crude) in animal feed and forages: near-infrared reflectance spectroscopic method. (989.03) Official Methods of Analysis. Association of Official Analytical Chemists. 15th ed.

Bartlett, M.S. 1937. Some examples of statistical methods of research in agricultural and applied biology (with discussion). J. Roy. Statist. Soc. Supp. 4:137–183.

Brownie, C., D.T. Bowman, and J.W. Burton. 1993. Estimating spatial variation of data from yield trials: A comparison of methods. Agron. J. 85:1244–1253.

Cullis, B.R., and A.C. Gleeson. 1991. Spatial analysis of field experiments– an extension to two dimensions. Biometrics 47:1449–1460.

Federer, W.T. 1956. Augmented (or Hoonuiaku) designs. Hawaiian Planters Record. 55:191–208.

Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447.

Johnson, R.A., and D.W. Wichern. 1998. Applied multivariate statistical analysis, Fourth ed. Prentice Hall, Inc., Upper Saddle River, NJ.

Littell, R.C., G.A. Milliken, W.W. Stroup, and R.D. Wolfinger. 1996. SAS system for mixed models. SAS Institute Inc., Cary, NC.

Marx, D.B., and W.W. Stroup. 1993. Analysis of spatial variability using PROC MIXED. *In* Proceedings of the 1993 Kansas State University Conference on Applied Statistics in Agriculture. Kansas State University, Manhattan, KS.

Miller, R.O., J. Kotuby-Amacher, and J. Rodriguez. 1998. Total nitrogen in botanical materials- automated combustion method. p. 116–117. *In* Western states laboratory proficiency testing program. soil and plant analytical methods. Version 4.10. Colorado State University, Ft. Collins. CO.

Padmore, J.M. 1990a. Fat (crude) or ether extract in animal feed, method 920.39. *In* K. Helrich (ed.) Official Methods of Analysis of the Association of Official Analytical Chemists, 15th ed. AOAC, Inc., Arlington, VA.

Padmore, J.M. 1990b. Minerals in animal feed–atomic absorption spectrophotometric method, method 968.08. *In* K. Helrich (ed.) Official Methods of Analysis of the Association of Official Analytical Chemists, 15th ed. AOAC, Inc., Arlington, VA.

Papadakis, J. 1937. Methods statistique por des experiences de champs'. Inst. Amel. Plantes, Salonique, (Greece). Bull Sci. 23.

Robinson, G.K. 1991. That BLUP is a good thing: The estimation of random effects. Statist. Sci. 6:15–51.

SAS Institute. 2003. SAS OnlineDoc., Version 9.1. SAS Institute, Cary, NC.

Searle, S.R. 1965. Computing formulae for analyzing augmented random complete block design. BU-207-M in the Biometrics Unit Mimeo Series. Cornell University.

Sebolai, B., J.F. Pedersen, D.B. Marx, and D.L. Boykin. 2005. Effect of grid size, control plot density, control plot arrangement, and assumption of random or fixed effects on non-replicated experiments for germplasm screening. Crop Sci. 45:.

Shenk, J.S., and M.O. Westerhaus. 1991. Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. Crop Sci. 31:469–474.

Steel, R.G.D. 1958. A class of augmented designs. Math Res. Center Tech Summary Report No. 56. University of Wisconsin.

USDA-ARS. 2005a. USDA, ARS, National Plant Germplasm System: Summary statistics of holdings as of 06 Feb 2005—Composition for Sorghum. http://www.ars-grin.gov/cgi-bin/npgs/html/stats/genus.pl?Sorghum; verified 22 November 2005.

USDA-ARS. 2005b. USDA, ARS, National Plant Germplasm System: Descriptors for sorghum. http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?69; verified 22 November 2005.

USDA-GIPSA. 1999. Official testing service for corn oil, protein and starch. 7 CFR Part 801, Federal Register Vol. 64, No. 74.

Ward Laboratories. 2005. Feed analyses: Wet chemistry. http://www.wardlab.com/; verified 22 November 2005.

Y S I incorporated. 2000. Determination of % cook in extruded cereal products using chemical solubilization. Application # 322. Yellow Springs, OH.

Zimmerman, D.L., and D.A. Harville. 1991. A random field approach to the analysis of field-plot experiments and other spatial experiments. Biometrics 47:223–239.