RESEARCH

Diversity of Maize Kernels from a Breeding Program for Protein Quality: I. Physical, Biochemical, Nutrient, and Color Traits

Abdullah A. Jaradat* and Walter Goldstein

ABSTRACT

It is a challenge to breed maize (*Zea mays* L.) varieties with enhanced levels of protein, oil, starch, and essential amino acid contents while maintaining competitive agronomic capability. We used multivariate statistical procedures to quantify total diversity and its components for 31 physical, biochemical, nutrient, and color traits of 1348 accessions from 13 maize populations. The germplasm belonged to stiff and non-stiff stalk heterotic groups at different stages of inbreeding and with opaque or translucent kernel endosperm. Cluster, principal components, and variance components analyses resulted in separating populations and traits into statistically different groups and helped to identify physical and color traits that may be useful for selecting accessions with high protein, amino acid, oil, and nutrient contents. We identified correlative relationships between groups of biochemical and nutrient traits linked to certain populations or physical or kernel color traits. Diversity analysis revealed variable relationships between and within total diversity and its components depending on kernel physical, biochemical, nutrient, and color traits. Sources of variation in and amount of variation attributed to total diversity for each of these traits were identified and quantified. Populations and genotypes within populations were identified as sources of variation for multiple kernel quality traits that can be used in developing high-quality maize varieties.

A.A. Jaradat, USDA-Agricultural Research Service, 803 Iowa Ave., Morris MN 56267; W. Goldstein, Mandaamin Institute, W2331 Kniep Road, Elkhorn WI 53121. Received 21 July 2012. *Corresponding author (abdullah.jaradat@ars.usda.gov; jarad001@umn.edu).

Abbreviations: a^*, redness–greenness; ANN, artificial neural network; b^*, yellowness–blueness; BSSS, Iowa Stiff Stalk Synthetic; GEM, Germplasm Enhancement of Maize; *Gst*, population differentiation; *Hs*, diversity within populations; *Ht*, total diversity; ISU, Iowa State University; L^*, darkness–lightness; MFAI, Michael Fields Agricultural Institute, NG, Nokomis Gold; NIRS, near-infrared spectroscopy; PC, principal component; PCA, principal components analysis; PLS, partial least squares, Q^2, validation coefficient of determination; UPGMA, unweighted pair group method with arithmetic average.

Maize provides about 30% of the food calories for approximately 4.5 billion people in 94 developing countries (Shiferaw et al., 2011); globally, however, 63% of maize is used for livestock feed. Maize kernel provides 365 kcal 100 g^{-1}; it is composed of approximately 73% starch, 10% protein, and 5% oil, with the remainder made up of fiber, vitamins, and macro- and micronutrients (Wang et al., 2012). The composition is largely controlled by the genetics of the endosperm sink, the maternal plant parent, and the environment (Nuss and Tanumihardjo, 2010). Since the 1920s successively released varieties having greater yield have had more starch and less protein contents (Scott and Blanco, 2009). However, kernel composition has been successfully manipulated by selection and may affect the nutritional value of the grain (Welch and Graham, 2004) and the value of end-use products (Ortiz-Monasterio et al., 2007). Farmers and the food and feed industries will continue to demand high-yielding maize varieties with enhanced nutritional and industrial qualities. Therefore, one of the most important

Published in Crop Sci. 53:956–976 (2013).
doi: 10.2135/cropsci2012.07.0437
© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

breeding objectives is to enhance the composition of maize grain as a source of energy, protein, oil, amino acids, and minerals for human food, animal feed, or industrial applications (Prasanna et al., 2001; Wang et al., 2012).

Although there are differences in physical and biochemical kernel traits among modern maize varieties, variation in these traits is relatively consistent within varieties as compared to variation among varieties and substantially smaller compared to variation found in landraces and open-pollinated varieties (Seebauer et al., 2010; Skogerson et al., 2010). Concentrations of quality traits, especially protein, are more diluted in kernels of modern high-yielding varieties as compared to those of landraces and old open-pollinated varieties (Fufa et al., 2003; Seebauer et al., 2010). The limited nutritional quality of maize protein, especially for monogastric animals, is attributed to deficiencies of a few essential amino acids (Flint-Garcia et al., 2009). Protein and oil contents usually are inversely related to each other and to starch content. Starch content is the least variable macromolecule in the maize kernel (CV = 2 to 3%) regardless of N application (Miao et al., 2006); this characteristic makes it difficult to manipulate the contents of macro- (protein and oil) and micromolecules (amino acids) and other quality traits (macro- and micronutrients). Starch types (amylose vs. amylopectine) can influence the physical characteristics and digestibility of maize kernels. High-oil maize (>6.0%), for example, can be a valuable livestock feed due to its high energy value and potential for replacing more expensive dietary sources of fats and proteins (Thomison et al., 2003). Although high protein content can reduce protein-rich additives in animal rations, its highly developed matrix within the endosperm may interfere with starch digestibility (Prandini et al., 2011).

Key physical, biochemical, and nutrient traits of the maize kernel, besides being critical factors that can affect the early development and vigor of maize seedlings, have been exploited by traditional (Salhuana et al., 1998) and advanced breeding and selection methods (Scott et al., 2008) in developing varieties with high food and feed nutritional quality. New genetic resources for quality traits and in-depth understanding of the genetic diversity and interrelationships between and within groups of quality and physical traits are becoming increasingly important as a result of genetic erosion of maize landraces and open-pollinated varieties (Jaradat et al., 2010). The use of such genetic resources to develop maize varieties with superior protein quality (Reeves and Cassady, 2002) and oil content (Thomison et al., 2003) were among the most important achievements of maize breeding research during the second half of the 20th century.

Several reports (Singh et al., 2011; Shiferaw et al., 2011) indicated that the majority of recently developed maize inbred lines in the United States and Canada are the products of breeding programs based on a highly stratified, narrow-based, and closed germplasm base. More heterogeneous and heterozygous maize varieties are needed to provide genetically broad-based populations as sources of quality traits (Taboada-Gaytan et al., 2010). Characterizing relevant germplasm pools, understanding the structure of their diversity, and elucidating interrelationships between and within physical–biochemical–nutrient–color traits should lead to enhanced germplasm use for the improvement of nutritional quality of maize (Harrigan et al., 2010; Mittelmann et al., 2011).

The germplasm comprising 1348 accessions derived from 13 populations in a breeding program to develop maize varieties with high protein quality and competitive agronomic traits was used in a comprehensive evaluation of 31 kernel-based traits. The objective of the study was to identify sources of variation, analyze genetic diversity, and quantify multivariate relationships between several physical, biochemical, nutrient, and color traits in the maize kernel.

MATERIALS AND METHODS

Germplasm Description

A breeding and selection program was initiated at Michael Fields Agricultural Institute (MFAI) (42°46′56.06″ N, 88°25′65″ W) in 1989 to develop modern open pollinated maize varieties. In 2001, the objective was modified to include developing high-quality maize varieties with enhanced levels of protein and the essential amino acid methionine, in addition to cysteine and lysine. The overall breeding strategy was to develop two separate heterotic groups with protein quality that may be used for producing high-yielding hybrids. The stiff stalk but not the non-stiff stalk heterotic group contained appreciable inheritance from the Iowa Stiff Stalk Synthetic (BSSS) population (Olsen et al., 2003). To represent the ongoing program, we sampled a set of 1348 breeding lines (accessions) from 13 populations grown in 2008 (Table 1). Seventy-six percent of the lines were obtained from nurseries planted in Elkhorn, WI, in 2008. Three of these nurseries from which 70% of the lines were obtained were sequential plantings on the MFAI farm on a Sebewa (typic Argiaquoll, fine, loamy over sandy skeletal, noncalcareous, mixed, mesic). Five percent of the lines were grown on a small nursery located 7 km south of that farm on McHenry (typic Hapludolf, fine, loamy, mixed, mesic) soil. Some 68% of the breeding lines from AR16021:B73 and CH05015-Mo17 and 38% of the lines from Nokomis Gold (NG) were grown in Ames, IA, at the Iowa State University (ISU) farm in 2008 by the USDA-ARS (L. Pollak, personal communication, 2008) on a Nicollet loam (fine loamy, mixed, superactive, mesic Aquic Hapludoll) with the rest grown at the MFAI nursery. Land used for all trials was organically certified with the exception of the small MFAI nursery, and organic management was applied to all nurseries in 2008. The whole germplasm was classified (Table 1) according to parentage, endosperm type, affinity to heterotic groups, and level of inbreeding or selfing (S1 to S4 indicate the number of selfed generations). Progressive selection for opaque kernels coupled with selfing led to an increase in the proportion of opaque kernels.

The germplasm was unbalanced as to the number of accessions within each combination of parentage, level of inbreeding, and endosperm type. There were 826 and 522 non-stiff stalk and stiff stalk genotypes, respectively, and 496 and 852 genotypes with opaque and translucent kernels, respectively. Number of

genotypes within populations ranged from three with opaque and three with translucent kernels in NG-Mo508-Mo506 to 117 with translucent kernels in DKXL370-S11. The genotypes DKXL370 and N11a were each part of the parentage of three of the crosses tested. Four parents (two male and two female) were involved in developing six populations; three parents (two female and one male) were involved in developing six more populations. The population NG was the only one with no additional parentage in its pedigree; it was developed from Native American maize varieties at MFAI. Although NG appears to combine well with stiff stalk and non-stiff stalk testers (J. Golden, personal communication, 2005) we classified it as a non-stiff stalk because it has no BSSS component in its background (Olsen et al., 2003). Some of these populations were derivatives of crosses made by the Germplasm Enhancement of Maize (GEM) project (www.public.iastate.edu/~usda-gem/), whose objective was to introgress the diversity associated with exotic maize into commercial, Corn Belt Dent, stiff stalk, and non-stiff stalk backgrounds.

Non-Stiff Stalk Germplasm

1895-DKXL212-N11a

The selection 1895 is an opaque, high methionine selection derived from the GEM program inbred DKXL370:N11-B-029-002 GEMN-0133 (USDA-ARS, 2013b). The GEMN-0133 inbred was derived at North Carolina State University from a cross between the Brazilian hybrid DKXL370 from the DeKalb Company with a non-stiff stalk inbred from a commercial company number "11" (M. Goodman, personal communication, 2009). The selection DKXL212:N11a is a high methionine, vitreous endosperm selection from the unreleased GEM program inbred DKXL212:N11a-139-1-1 (USDA-ARS, 2013b). The latter inbred was developed at the University of Delaware from a cross between a Brazilian inbred DKXL212 from the DeKalb company with a non-stiff stalk inbred "a" from commercial company number "11" (J. Hawk, personal communication, 2009). Early segregates from the cross between 1895 and DKXL212:N11a-139-1-1 were sib mated for two seasons and self-pollinated in the third season; the opaque phenotype in this population is associated with reduction in kernel weight.

CH05015-Mo17

This is a family of derivatives of the first backcross between an accession of the Chilean Cateto landrace Camelia CHZM 05 015 (PI 467165) with the inbred Mo17 as the recurrent parent. The original Chilean accession is characterized as having good agronomic traits and relatively good combining ability for yield (Salhuana et al., 1998) and has a bright orange color (USDA-ARS, 2013c) that is probably associated with a high carotenoid content (Kuhnen et al., 2011).

GQ-N16-N12

This is a family of derivatives directly derived from an unreleased F_1 hybrid from the GEM program called Gold Queen N1612. The cross between an older Corn Belt landrace called Gold Queen and a non-stiff stalk inbred from commercial company number "16" was subsequently crossed with another non-stiff stalk inbred from company number "12." Lines derived from this cross showed high protein and essential amino acid contents.

Table 1. Maize germplasm classification according to parentage, endosperm type, heterotic group, and level of inbreeding or selfing stage (S1 to S4).

Population	Heterotic group†	First parent‡		Second parent		Inbreeding or selfing stage										Total
						S1 (1 yr)		S2 (2 yr)		S3 (3 yr)		S4 (4 yr)		Open pollinated		
						Opaque	Translucent	Opaque	Translucent	Opaque	Translucent	Opaque	Translucent	Opaque	Translucent	
						Number of genotypes (10 kernels and/or genotype)										
1895-DK212-N11a	NSS	DK370	N11	DK212	N11a	22	47									69
CH05015-Mo17	NSS	CH05015	Mo17		Mo17	9	21			20	32					82
DKXL370-N11a-N20	NSS	DKXL370	N11a		N20	36	54									90
DKXL888-N11a-N17	NSS	DKXL888	N11a		N17	41	72									113
GQ-N16-N12	NSS	GQ	N16		N12	5	29	4	10							48
LH82-AR16021	NSS	LH82	AR16021							37	112					149
Mo42-N220A	NSS	Mo42	N220A			25										25
NG-Mo508-Mo506	NSS	NG‡		Mo508	Mo506	3	3									6
NG	NSS	NG	NG			21	44	77	58	3	19	14	8			244
AR16021-B73	SS	AR16021	B73			9	47	15	8	72	43			17	19	230
CHIS740-S14-S12	SS	CHIS740	S14		S12	32	70									102
DKXL370-S11	SS	DKXL370	S11			7	117									124
LH119-AR16035	SS	LH119	AR16035			27	39									66
Total						237	543	96	76	132	206	14	8	17	19	1348

†NSS, non-stiff stalk; SS, stiff stalk.
‡NG, Nokomis Gold.

DKXL370-N11a-N20

This is a family of derivatives directly derived from an unreleased F_1 hybrid from the GEM program. The cross was between a Brazilian hybrid from the DeKalb company (DKXL370) with a non-stiff stalk inbred "a" from commercial company number "11," which was subsequently crossed with another non-stiff stalk inbred from company number "20." Lines derived from this cross showed high protein and essential amino acid contents.

DKXL888-N11a-N17

This is a family of derivatives directly derived from an unreleased F_1 hybrid from the GEM program. The cross was between a Thai hybrid from the DeKalb company (DKXL888) with a non-stiff stalk inbred "a" from a commercial company number "11," which was subsequently crossed with another non-stiff stalk inbred from company number "17." Lines derived from this cross showed high protein and essential amino acid contents.

LH82-AR16021

The inbred LH82 was developed by the Holden's Seed Company (PI 601170). Its pedigree is 610 × LH7 (LH7 is a derivative of W153R, which is a selection from the population Minnesota 13 [Gerdes et al., 1993]) (Plant Variety Protection certificate in 1984 for LH82 #8500037; USDA-AMS, 2013b). The AR16021 is a family of selections derived from the Argentinean accession ARZM 16 021 (PI 516022; landrace Cristalino Colorado). This accession was characterized by the Latin American Maize Project as having good yield and combining ability with testers (Salhuana et al., 1998). It was subjected to four growing seasons of half-sib selection at MFAI with primary selection practiced for adaptation (earliness, per se yield, disease resistance, and stalk and root strength). Pollen from a bulk of the resulting lines was crossed with LH82._The genetic source of the opaque trait is probably associated with the Argentinean accession as occasional floury kernels were found in the population during the half-sib selection process.

Mo42-N220A

The Mo42 (Ames 20120) is a third cycle derivative of the cross between Mo17 × C103 (USDA-ARS, 2013a), which are inbreds belonging largely to the Lancaster heterotic group. The N220A is a derivative of Mo17 from the University of Nebraska Research Station. In this study, we tested S1 lines derived from F_2 segregates of the cross. The opaque trait was observed in Mo42 and in the offspring of this cross.

Nokomis Gold

Nokomis Gold is a population was derived from Hopi Flour maize obtained from the Hopi Indian Reservation, Hopi, AZ (35°54′17.58″ N, 110°37′00.81″ W), with smaller contributions from Zuni Blue maize (PI 213799) and from crosses between Navajo Eagle maize (PI 222285) with Aren's synthetic (a prolific synthetic developed by the late Albert Arens of Hartington, NE). The population NG was released for use by farmers in 1998, but selection continued since then at MFAI. The initial half-sib selection program emphasized stalk and root strength and grain yield but in later years the program shifted to inbred line development to combine agronomic traits and grain kernel quality.

NG-Mo508-Mo506

The Mo506 and Mo508 are white maize lines developed at USDA-ARS, Columbia, MO (L. Darrah, personal communication, 2005). These lines are backcross introgressions of the gene *Ga1S* into K6 and K30, respectively (Gerdes et al., 1993). The inbred K6 was derived from Pride of Saline at Kansas State University. The line H30 was derived from the cross (K64 × 38-11)38-11. The inbred K64 was derived from the population Pride of Saline. The inbred 38-11 was derived from the outcross Funk 176A, which was derived from a Reid Yellow Dent population. These lines represent older Corn Belt inbreds. The cross between Mo506 and Mo508 was donated by Cornell University to MFAI (F. Kutka, personal communication, 2005); it was maintained at Cornell as a heterogeneous donor of *Ga1S* and was used as a male parent in crosses with NG.

Stiff Stalk Germplasm

AR16021-B73

This is a family of derivatives of the cross between an accession of the Argentinean 'Cateto' flint ARZM 16 021 (PI 516022; landrace Cristalino Colorado) with the inbred B73. The original F_1 seed was donated to MFAI from the GEM program in 2001. Initially it was bred using half-sib selection for per se earliness, other agronomic traits, and grain quality at MFAI with periodic recombination of the better lines. Cooperative yield trials with USDA-ARS (L. Pollak, personal communication, 2007) indicated that lines of AR16021-B73 had a generally high level of combining ability when crossed with commercial non-stiff stalk testers.

CHIS740-S14-S12

This is a family of derivatives directly derived from an unreleased F_1 hybrid from the GEM program called CHIS740:S1412. The cross was made between Mexican landrace CHIS740 from Chiapas (PI 583890) identified as a derivative of the landrace Tuxpeno and a stiff stalk inbred from commercial company "14," which was subsequently crossed with another non-stiff stalk inbred from company number "12." The Mexican landrace has white translucent kernels and high combining ability for yield. The seed derived from this cross was selected for high carotenoid content. Lines were subsequently selected for high protein, essential amino acid contents, and opaque kernels, the genetic source of which is unknown.

DKXL370-S11

This is a family of derivatives from a single ear selection from an S3 grow-out of a GEM breeding line GEMS-0029 (PI 639054). The GEM inbred was derived by Major Goodman at North Carolina State University from a cross between a Brazilian hybrid DKXL370 from the DeKalb Company with a stiff stalk inbred from a commercial company number "11." A single S1 selection was characterized as having a high carotenoid and methionine content. Breeding lines derived from this selection were further selected for earliness, good agronomic traits, and grain quality with several cycles of inbreeding followed by recombination. The genetic source of the opaque trait is unknown.

LH119-AR16035

The inbred LH119 was developed by the Holden's Seed Company (PI 600954). Its pedigree is B73 (2x) × H93 (H93 is a derivative of B37) (Plant Variety Protection certificate in 1983 for LH119 #8200064; USDA-AMS, 2013a). The AR 16035 is a family of selections derived from the Argentinean Cateto flint accession ARZM 16 035 (PI 516036; landrace Cristalino Colorado). This landrace produced higher yields than any other landrace (Salhuana et al., 1998) when tested in hybrid combination with stiff stalk and non-stiff stalk testers in Iowa. Since 2001 it underwent repeated cycles of half sib selection and recombination for earliness and per se agronomic performance at MFAI.

Kernel Physical Measurements

Ten kernels each of 1348 germplasm accessions were taken at random from a larger sample and their weight was recorded before further analyses. Digital imagery and analysis procedures (Rasband, 2007; Igathinathane et al., 2008) were used to capture and analyze morphological traits of individual maize kernels (i.e., area, perimeter, width, length, major axis, minor axis, and circularity). The 10 kernels per accession and a scale (1.0 cm in length) were manually positioned on a platform to ensure that kernels were totally separated and their embryos were facing up. They were then digitally photographed using a Nikon D70S digital camera with 1504 by 1000 pixel resolutions. The number of kernels per image was verified by the number of objects generated by the ImageJ software program (Igathinathane et al., 2008). Kernel density (i.e., mass of kernels in a defined volume) was measured using a pycnometer. The lower densities were considered to be softer (Fox and Manley, 2009). The method depends on moisture content of the kernels; therefore, densities were estimated at similar moisture contents.

Kernel Biochemical Analyses

The near-infrared spectroscopy (NIRS) analyses reported in this paper resulted from analyzing 1348 kernel samples at MFAI with the Bruins OmegAnalyzerG (Bruins Instruments) equipped with calibrations for moisture, protein, oil, starch, methionine, lysine, and cysteine and/or cystine from the ISU Grain Testing Laboratory. The technique provides a basis for comparison of an unknown sample to a well-characterized control. In analyzing these macromolecules, the calibrations are based on multivariate and multiple wavelength calibration technique of chemometrics (Osborne, 2001; Hacisalihoglu et al., 2010). In this procedure, prediction models are developed for each macromolecule as dependent variable while the reduced spectral data are used as independent variables. This step is performed after principal components analysis (PCA) is performed on the spectral data.

Over a 2-yr period, kernel samples from the MFAI research program on high-protein maize were scanned with NIRS technology and spectra were collected and NIRS-based calibrations were developed at the ISU Grain Testing Laboratory, Ames, IA. In each year, 100 of the most spectrally diverse samples were then sent to University of Missouri Agriculture Experiment Station Chemical Laboratory to determine amino acid contents using high pressure liquid chromatography. Prediction and validation models for the wet chemistry and spectral data were developed on the basis of artificial neural network (ANN) and partial least squares (PLS) option in the nonlinear iterative partial least squares (NIPALS) module in a multivariate software program (The Unscrambler software, version 10.1; CAMO, 2011). Cross-validation in PLS was performed by removing each sample from the calibration dataset, and a model was then calibrated on the remaining data points. The value for each removed sample was predicted, and the prediction residual was estimated. The validation residual variance and root mean square error of the cross-validation process were calculated after combining all prediction residuals. The standard error of calibration, the coefficient of determination in calibration, and the root mean square error of cross validation were used in evaluating the predictive ability and goodness of fit of the models.

The NIRS-based validation models estimated cysteine, lysine, and methionine with validation coefficient of determination (Q^2) values of 0.79, 0.84, and 0.75, respectively, as compared with the respective values of 0.80, 0.54, and 0.39 without NIRS calibration. The resulting calibration model was licensed by MFAI to the Bruins Manufacturing Company (Salem, NH).

Kernel Nutrients Analyses

Kernel samples were dried at 45°C in a forced air oven until no further reduction in weight occurred. Kernels were ground and placed through a 1-mm screen (Thomas Scientific) and then one subsample was used to determine C and N and another to determine macro- and micronutrients. Carbon and N were determined using a LECO FP-428 analyzer and then the C:N ratio was calculated. Digestion of kernel ground material for macro- and micronutrient (Ca, Cu, Fe, K, Mg, Mn, P, S, and Zn) determination followed the procedure outlined in the USEPA 5051 method (USEPA, 2007). We adapted this procedure using the Mars Xpress Microwave System from CEM Corp. sample preparation note XprAg-1. The procedure uses 55-mL Teflon tubes in a 40-unit carousel. A 0.5-g sample weight was digested with 6.5 mL HNO_3 (70% trace material analysis) using a 15 min ramp program set to a power maximum of 1200 W and held for 15 min. The samples then were allowed to cool to room temperature and then transferred to 50-mL volumetric flasks and taken to volume with Milli-Q water (Millipore Corp.). Chemical analyses were completed using the Varian Vista-Pro CCD (charged coupled device) (Varian Inc.) Simultaneous ICP–OES (inductively coupled plasma–optical emission spectroscopy) instrument. The elemental standards "MNUSDA-STD 1A" and "MNUSDA-STD 2 were prepared according to Inorganic Ventures, Lakewood, NJ.

Kernel Color Analyses

The three-dimensional red–green–blue color space of each of 1348 samples were used to estimate their $L^\star a^\star b^\star$ color space (darkness–lightness [$L^\star$], redness–greenness [$a^\star$], and yellowness–blueness [$b^\star$] descriptors according to Commission International de l'Eclairage recommendations [CIE, 2007; Darrigues et al., 2008]).

Statistical Analyses

Hierarchical Cluster Analysis

Separate and two-way joint hierarchical clustering of 31 kernel traits and 13 maize populations were performed using Euclidean distance and the unweighted pair group method with arithmetic

average (UPGMA) linkage clustering procedure (StatSoft, 2011). The two-way joint data was transformed to develop a ±1.0 scale, which then was used to color code the joint clustering of populations and kernel traits (Fig. 1) and to identify populations with preferred levels of single or multiple traits (Cañas et al., 2011).

Principal Components Analysis

We used PCA on the standardized means of quantitative traits as a linear dimensionality reduction technique to identify orthogonal directions of maximum variance (at the prediction [R^2] and validation [Q^2] phases of model building) in the original data set, including populations, classification criteria (i.e., endosperm type, heterotic groups, and inbreeding stage), and kernel physical, biochemical, nutrient, and color data. The data was projected into the first principal component (PC1) and PC2 orthogonal dimensions. Correlations between the initial traits and each of these PCs (i.e., loadings) were used to interpret results of the analysis (Payne et al., 2007). Separate PCAs were performed for each group of traits (i.e., kernel physical, biochemical, and nutrient traits) and then we estimated the correlation coefficients between individual traits in the remaining groups and each PC.

Variance Components Analyses

Percent variance and its level of significance (p) due to main effects of five sources of variation (i.e., populations, genotypes within populations, endosperm type, heterotic groups, and inbreeding stage), least squares mean, and coefficient of variation (CV %) for kernel physical, biochemical, nutrient, and color traits were estimated using GenStat version 10.1 (Payne et al., 2007). Single degree of freedom contrasts (in the case of endosperm types and heterotic groups) or mean separation (in the case of inbreeding stage) at the 5% level of probability were used to identify significant differences between categories of each source of variation.

Diversity Analysis

The mean and SD calculated for each of 31 kernel physical, biochemical, nutrient, and color traits listed in Fig. 1 were used to classify each accession into one of three discrete categories: low, medium, or high. If a trait value was equal to or 1 SD less than the mean, it was classified as "low"; if it was equal to or higher than the mean plus 1 SD, it was classified as "high"; otherwise it was classified as "medium" (Zar, 1996). The frequency data was used according to Yeh et al. (2000) in estimating total diversity (Ht) and its components, that is, diversity within populations (Hs) and population differentiation (Gst). Statistical moments and correlation and regression coefficients describing the relationships between these diversity components and their level of significance were estimated (StatSoft, 2011).

RESULTS

Cluster Analyses

The UPGMA hierarchical clustering procedure totally separated the 13 maize populations into two groups, each of which displayed Euclidean distances between 15 and 45% (Fig. 1). The first group was composed of one non-stiff stalk and three stiff stalk populations. The second group was composed of eight non-stiff stalk and one stiff stalk populations. In this group, two non-stiff stalk populations (DKXL370-N11a-N20 and NG-Mo508-Mo506) were totally separated from the rest. The UPGMA hierarchical clustering procedure also separated starch and C:N from the remaining kernel traits, which were separated into two groups at the 60% Euclidian distance. Eight nutrients (N, S, Mg, Mn, P, C, Fe, and Zn) and all three essential amino acids (lysine, cysteine, and methionine) were closely clustered with protein. Oil, in addition to three nutrients (K, Ca, and Cu), clustered with the kernel physical and color traits. However, these traits can be separated into three groups at about 40% Euclidean distance.

The joint clustering procedure indicated that four populations (1895-DK212-N11a, AR16021-B73, LH119-AR16035, and DKXL370-S11) (Fig. 1) differed from the remaining populations by having lower than average content of starch, C:N, lysine, Mg, Mn, P, Fe, and Zn and larger values for kernel moisture and kernel density. Population–trait groupings can be identified on the basis of the color-coded joint clustering procedure. Two populations (LH82-AR16021 and Mo42-N220A) have above mean values for oil, two color descriptors ($L^\star$ and $b^\star$), two nutrients (Ca and Cu), and five (weight, area, width, major axis, and circularity) kernel physical traits. On the other hand, a single population (NG-Mo508-Mo506) had below average N, protein, and cysteine but above average methionine and S, and all populations in this group had kernels with below average minor axis, with one exception, and heavier kernels than the remaining populations.

Principal Components Analyses

Calibration (R^2) and validation (Q^2) variances explained by the PC1 and PC2 derived from the 13 maize populations, three categorical descriptors (endosperm type, heterotic groups, and inbreeding status), and 31 kernel physical, biochemical, chemical, and color traits measured on kernel samples of maize populations are presented in Fig. 2. At the calibration phase, PC1 and PC2 explained 0.47 and 0.23 of total variation, respectively, and 0.36 and 0.18 of total variation, respectively, at the validation phase of model building. Populations were separated into four groups (I–IV) along PC1 and PC2, with large positive (AR16021-B73) or negative (CH05015-Mo17) loadings or small positive (GQ-N16-N12) or negative (NG) loadings on PC1, which explained 50% more variation than PC2 at the validation stage of model building. The population NG-Mo508-Mo506 differed from the rest by having negative and positive loading on PC1 and PC2, respectively. The inbreeding stage and heterotic groups contributed the least amount of variation to both PCs followed by endosperm types. Calcium and Cu had the smallest loadings as compared with the remaining nutrients. The trait–population associations in each of the four quadrants (I–IV) (Fig. 2) suggested that populations in quadrants I and II, as compared to those in quadrants III and IV, have larger kernels, mostly with translucent endosperm,

and higher starch, C:N, and oil contents. Conversely, populations in quadrants III and IV have smaller kernels, mostly with opaque endosperm, and larger protein, amino acids, and nutrient contents.

Each of $L^\star$, $a^\star$, and $b^\star$ had unique loadings on both PCs (Fig. 2). The dark–light ($L^\star$) descriptor had large positive and negative loadings on PC1 and PC2, respectively, and the red–green ($a^\star$) descriptor had relatively smaller negative and larger positive loading on PC1 and PC2, respectively, while the yellow–blue ($b^\star$) descriptor had intermediate positive loadings on both PCs. The positive associations between $L^\star$ and $b^\star$ on PC1 and between $a^\star$ and $b^\star$ on PC2 suggest that the kernel color patterns are complex and singly or in combination are disassociated from or associated with different populations and with different levels of kernel physical, nutrient, and biochemical traits.

Significant ($p < 0.05$) cumulative validation variance (Q^2), loadings (underscored) on and significant ($p < 0.05$) correlation coefficients (in *italics*) with the first three PCs derived from kernel physical, biochemical, nutrient, and color traits of maize populations are presented in Table 2. All listed loadings (i.e., correlation coefficients between traits and their respective PC) and correlation coefficients between traits and PCs, other than the ones they contribute to, are significant at the 5% level of probability. Three PCs explained cumulative 0.77, 0.86, and 0.62 of total variation at the validation stage in kernel physical, biochemical, and nutrient traits, respectively. The first PC in each case explained more than 50% of that variation and displayed positive loadings of kernel physical and nutrient traits on their respective PC1 and positive loadings of starch and C:N on PC1 of biochemical traits. Different trait combinations contributed to additional 0.15, 0.18, and 0.11 of variation explained by PC2 in kernel physical, biochemical, and nutrient traits, respectively, and the respective contribution of PC3 were 0.09, 0.07, and 0.07. Each PC described one or more characteristics of kernel physical, biochemical, or nutrient traits.

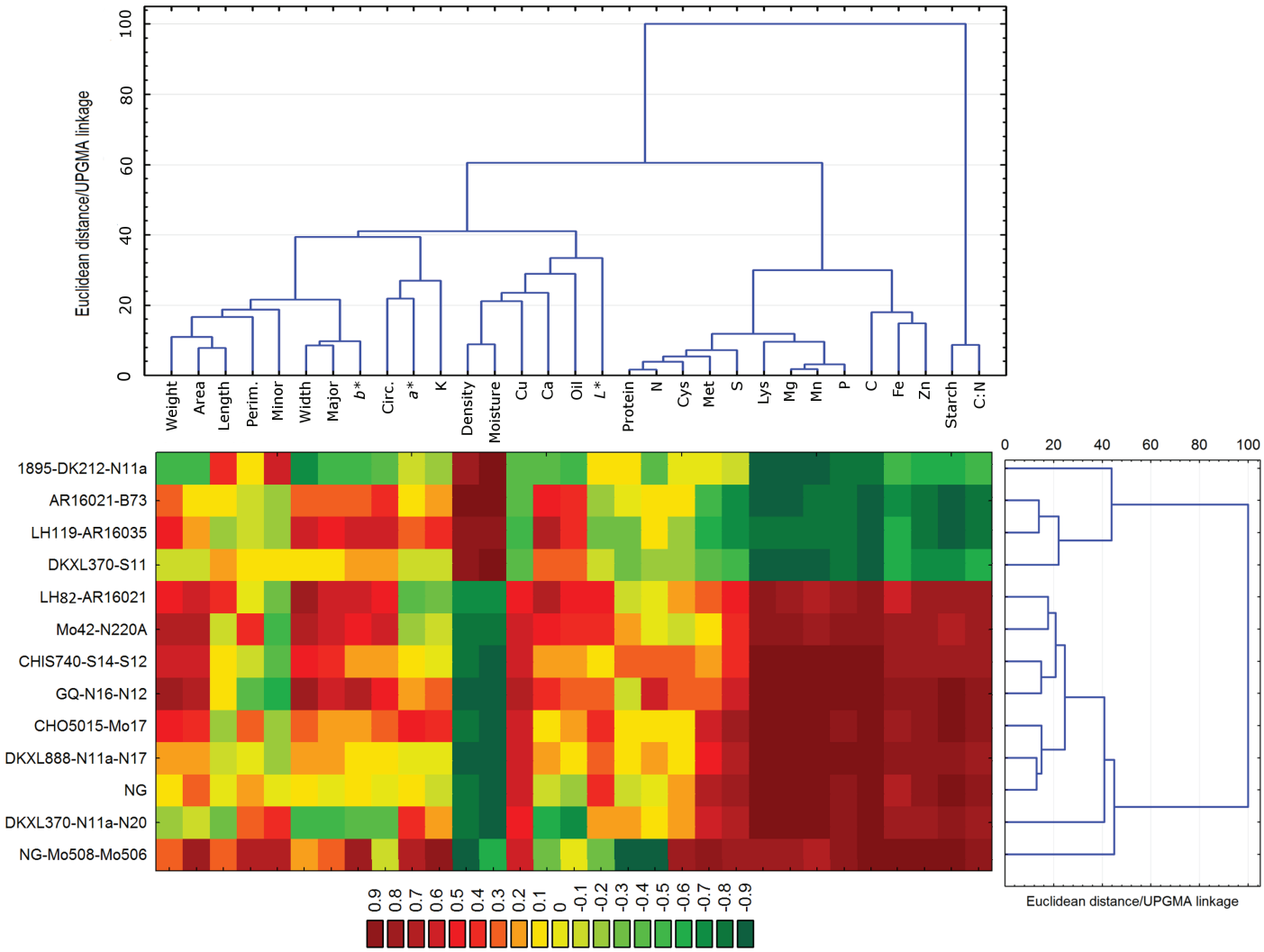

Figure 1. Joint and hierarchical clustering of 31 traits measured on kernels of 13 maize populations and expressed as standard deviation units from a zero mean and unit variance for each trait using Euclidean distance and an unweighted pair group method with arithmetic average (UPGMA) linkage clustering procedure. Perim., perimeter; b^*, yellowness–blueness; Circ., circularity; a^*, redness–greenness; L^*, darkness–lightness; NG, Nokomis Gold. Perim. and circ. are estimated as $4\pi \times$ [kernel area/(perimeter)^2].

None of the kernel physical traits were correlated with the first PC derived from biochemical traits (PC1 derived from biochemical traits). Kernel weight and density were the only traits negatively correlated with PC2 derived from biochemical traits, and all kernel physical traits (except circularity) were positively correlated with PC3 derived from biochemical traits. Kernel moisture was the only biochemical trait correlated with PC1 derived from physical traits; different combinations of biochemical traits were positively and negatively correlated with PC2 and PC3 derived from physical traits. Protein, cysteine, lysine, and methionine were positively and negatively correlated with the PC1 and PC2 derived from nutrients, respectively. Starch and C:N exhibited opposite correlations with PCs derived from nutrients, and none of the nutrients was correlated with PC3, which was only defined by calcium and explained 0.07 of total variation in nutrient composition of maize kernels. Potassium was the only nutrient negatively correlated with PC1 and PC2 derived from physical traits and positively correlated with PC2 derived from biochemical traits. The remaining nutrients, except Ca, Cu, and Fe, were negatively correlated with PC1 or with PC3 derived from biochemical traits. The three color space descriptors contributed to both PCs (Fig. 2) and displayed different positive or negative correlations with PC1 and PC2 derived from kernel physical and biochemical traits but not with PCs derived from nutrient traits.

Variance Components Analyses

A wide range of total percent variance, explained by the main effects of all five variance components, was found for kernel physical traits (Table 3). Differences between endosperm types contributed the most to the smallest variance (23.0% for kernel circularity) and differences between populations contributed the most to the largest (84.2% for kernel density) total variance. Differences between populations but not between genotypes within populations explained the largest portions of variances in all kernel physical traits. The level of variation, at the population level, expressed as a percentage of the mean (i.e., CV) ranged from 4.1% for kernel density to 18.5% for kernel weight.

Means of the kernel physical traits of opaque and translucent endosperm types differed significantly, with the translucent type having significantly larger values of all traits except kernel circularity. However, variances explained by these differences were generally small and largely nonsignificant, except for kernel weight and kernel density. Percent variances due to differences between stiff and non-stiff stalk heterotic groups were small in magnitude (<3.9%) with the only significant mean differences found for kernel weight, which displayed a larger CV value for stiff stalk (21.2%) than for the non-stiff stalk (16.2%) heterotic groups. Percent variance explained by differences between inbreeding stage (including selfing and open-pollinated germplasm) were all significant. They constituted the second largest variances (10.5– 29.4%) after those explained by differences between populations. Kernel density was an exception where variances explained by differences between endosperm types (33.2%) and between genotypes within population (27.2%) had the largest variances. The open-pollinated or S4 germplasm had significantly larger mean values for all kernel physical traits, except kernel density (open-pollinated) and circularity (S4).

The smallest (57.9%) and the largest (90.2%) total variances in biochemical traits were found for C:N and kernel moisture content (Table 4). Differences between genotypes and differences between endosperm types contributed the most to total variance of the C:N (28.8%) and kernel moisture (31.9%), respectively. Variances due to differences between populations were significant but were not the largest (except for protein; 12.8%) as compared to those due to differences between endosperm types (kernel moisture, lysine, and oil content) or to those due to inbreeding stage (methionine, cysteine, starch, and C:N). Variances due to differences between genotypes within populations were either marginally significant (kernel moisture, $p = 0.08$, and lysine, $p = 0.08$) or not significant (remaining traits); however, a few of these variances (e.g., kernel moisture, 22.5%, and lysine, 33.8%) were large in magnitude. Although both endosperm types did not differ significantly in their protein content, kernels with opaque endosperm had significantly larger mean values of lysine, methionine, oil, and C:N than translucent types while the latter had significantly larger mean values for kernel moisture, cysteine, and starch content.

Variances due to heterotic groups were very small (<3.2%) and nonsignificant with only significant differences in kernel moisture and lysine content. Variances due to the level of inbreeding were all significant and resulted in significant mean differences for all biochemical traits. The advanced inbreeding stage (S4) had significantly larger mean values for kernel moisture, protein, cysteine, lysine, and methionine. With the exception of starch content (CV = 2.89%), the remaining biochemical traits displayed small to relatively large CV values (9.46–26.5%) at the population level; these values were similar in magnitude to those displayed by the remaining sources of variation except inbreeding stage. Small differences between CV values at different inbreeding stages were found for kernel moisture content, cysteine, lysine, methionine, and oil content but larger differences were found for protein and C:N.

Six nutrients (C, K, Mg, Mn, P, and S) had more than 50% of their total variances explained by all five main sources of variation (Table 5). Calcium had the smallest total variance (19.7%) and K the largest (62.3%) whereas the remaining nutrients had intermediate values. Differences between populations contributed the most to total variances in Ca, Fe, K, P, and Zn while differences due to level of inbreeding contributed the most, in a

decreasing order, to variances in S, Mg, Mn, P, Fe, and N. At the population level, extremely small and large values were found for C (0.72%) and Cu (82.3%). However, except for Ca (CV = 40.4%), the remaining CV values ranged from 10.8% for Mg to 27.7% for Fe. Variances due to differences between genotypes within populations were significant for C, Mg, P, and S, marginally significant for N, Ca, and Mn, and nonsignificant for Cu, Fe, K, and Zn. Variances due to endosperm types were small (<3.6%), except for C (4.2%; p = 0.06) and K (12.1%; p = 0.02). Kernels with opaque endosperm had significantly larger C, N, K, Mg, Mn, and P contents, but there were no significant differences between opaque and translucent endosperm types for the remaining nutrients.

All variances attributed to differences between heterotic groups were extremely small (<3.2%) and nonsignificant. Similarly, differences between their respective means of stiff stalk and non-stiff stalk heterotic groups for all nutrients were not significant and their CV values were comparable to those found for populations. Variances due to level of inbreeding were relatively large and significant for all nutrients except Ca; likewise, significant differences for different levels of inbreeding were found between mean values of all nutrients, except Ca. Largest nutrient contents for different nutrients were mostly found at the advanced inbreeding stages (e.g., S3 for Cu, Fe, and K and S4 for N, Mg, Mn, and S).

Total variances for kernel color space descriptors ($L^\star$, 94.3%, $a^\star$, 82.4%, and $b^\star$, 90.7%) were larger than variances for kernel physical, biochemical, and nutrient traits (Table 6). At the population level, the $L^\star$ descriptor displayed the smallest (CV = 3.2%), $b^\star$ displayed an intermediate (CV = 10.9%), and $a^\star$ the largest (CV = 16.07%) level of variation. Differences between inbreeding stages contributed the largest to total variance in $L^\star$ (32.1%) whereas differences between endosperm types contributed the largest to variances in $a^\star$ (39.5%) and $b^\star$ (46.5%). Variances due to differences between populations for $L^\star$ (25.9%), $a^\star$ (19.7%), and $b^\star$ (18.4%) ranked second and were all significant. Genotypes within populations displayed a marginally significant difference for $a^\star$ (p = 0.06) but not for $L^\star$ or $b^\star$. Variances due to differences between endosperm types were significant for all three kernel color space descriptors; their means were highly significant, with the translucent

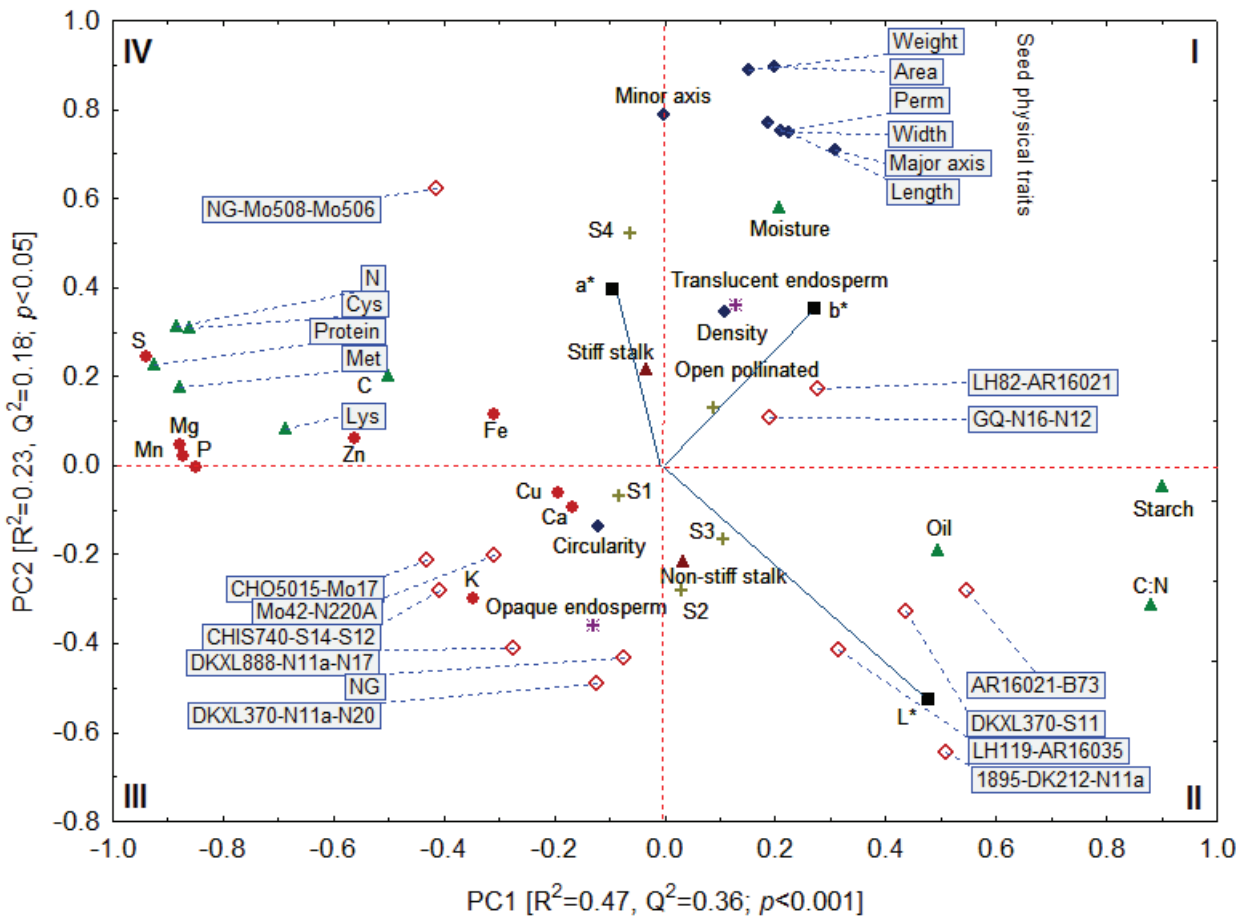

Figure 2. Calibration (R^2) and validation coefficient of determination (Q^2) variances explained by the first principal component (PC1) and PC2 derived from 13 maize populations, three categorical descriptors (endosperm types, heterotic groups, and inbreeding status), and 31 kernel physical biochemical, chemical, and color traits measured on kernel samples of maize populations. a^*, redness–greenness; b^*, yellowness–blueness; L^*, darkness–lightness; NG, Nokomis Gold. S1 to S4 indicate the number of selfed generations.

endosperm type having larger values (more negative $a^\star$ values) than the opaque endosperm type.

There were no significant differences between mean $L^\star$ and $b^\star$ values due to heterotic groups. However, the non-stiff heterotic group had significantly larger (i.e., more negative) $a^\star$ mean value as compared to the stiff stalk heterotic group. Nevertheless, variances due to differences between these groups were small (<5.6%) and nonsignificant. Variances due to inbreeding stage were marginally significant and their means displayed significant differences between two inbreeding groups for each of $L^\star$ and $a^\star$ and between three inbreeding groups for $b^\star$.

Diversity Analyses

The statistical moments, correlation coefficients, and frequency distribution of *Ht*, *Hs*, and *Gst* coefficient (Fig. 3) suggested that a large (*Hs* = 0.63) portion of *Ht* is partitioned within populations while the *Gst* coefficient comprised a relatively small portion of that variation. Both *Ht* and *Hs* were negatively skewed, *Gst* was positively skewed, and all three variables had positive kurtosis. The overall correlation coefficients between *Ht* and each of *Hs* ($r = 0.53$, $p = 0.01$) and *Gst* ($r = 0.36$, $p = 0.01$); although significant, they were not exceptionally high while the *Gst* coefficient was negatively and significantly ($r = -0.60$, $p = 0.01$) correlated with *Hs*.

Total diversity and *Hs* for each of kernel physical, biochemical, nutrient, and color traits exhibited wide ranges of values and linear relationships depending on the main source of variation used to calculate these statistics (Fig. 4). The largest ranges in variation of *Ht* for kernel physical, biochemical, nutrient, and color traits were associated with the level of inbreeding. For kernel physical traits the widest ranges for *Ht* and *Hs* were associated with level of inbreeding followed by heterotic groups and endosperm type. Total diversity for physical traits based on endosperm type was marginally correlated with *Hs* ($r = 0.42$, $p = 0.08$), and the latter can be estimated with small certainty ($R^2 = 0.18$).

Stronger relationships were found between *Ht* and *Hs* of biochemical traits as compared to kernel physical

Table 2. Significant ($P < 0.05$) cumulative validation coefficient of determination (Q^2), loadings (underscored) on, and significant ($p < 0.05$) correlation coefficients (in *italics*) with the first three principal components (PCs) derived from kernel physical, biochemical, nutrient, and color traits of maize populations.

		Physical			Biochemical			Nutrients		
		PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
Traits	Q^2	0.53	0.62	0.77	0.61	0.79	0.86	0.44	0.51	0.62
	Trait	Loadings or *correlation coefficients*								
Physical	Weight	0.87	0.39	0.20		−0.43	0.42			
	Density		0.87	−0.46		−0.77	0.25			
	Area	0.95					0.33			
	Perimeter	0.87		−0.30			0.24			
	Width	0.87	−0.26				0.27			
	Length	0.88					0.24			
	Major axis	0.86	−0.40	−0.23			0.25			
	Minor axis	0.79	0.40	0.41			0.30			
	Circularity			0.67						
Biochemical	Moisture	0.38	0.65	−0.29		−0.83	0.55			
	Protein			*0.21*	−0.98			0.75	−0.46	
	Lys		*−0.24*	*0.39*	−0.84	0.37		0.46	−0.32	
	Met			*0.31*	−0.98			0.67	−0.43	
	Cys		0.31		−0.92	−0.38		0.71	−0.45	
	Oil		−0.47	0.20		0.87	0.46			
	Starch			*−0.33*	0.94	−0.30		−0.72	0.41	
	C:N		−0.24		0.94	0.29		−0.73	0.47	
Nutrients	N		0.24		−0.94	−0.28		0.74	−0.48	
	C				−0.48	*0.23*	*0.53*	0.47	−0.40	
	Ca									0.96
	Cu							0.25	0.63	
	Fe							0.49	0.58	
	K	−0.21	−0.32			*0.27*	*−0.60*	0.54	0.24	
	Mg				−0.68		*−0.30*	0.95		
	Mn				−0.67		−0.24	0.99		
	P				−0.65		*−0.33*	0.93		
	S				−0.86			0.93	−0.25	
	Zn				−0.35			0.77	0.50	
Color†	L*	−0.69					−0.33			
	a*	0.24	0.28			−0.31	0.27			
	b*	0.63	−0.30							

†*L**, darkness–lightness; *a**, redness–greenness; *b**, yellowness–blueness.

traits. The strongest relationship between *Ht* and *Hs* was found for endosperm type ($r = 0.87$, $p = 0.0001$) where *Hs* can be estimated with high certainty ($R^2 = 0.75$). Although the correlations between *Ht* and *Hs* based on inbreeding stage ($r = 0.41$, $p = 0.009$) and heterotic group ($r = 0.47$, $p = 0.06$) were significant, the level of certainty in estimating *Hs* in these two cases remained very small ($R^2 = 0.16$ and 0.22, respectively).

The strongest and highly significant correlations ($p = 0.0001$) were found between *Ht* and *Hs* for nutrient content for all three sources of variation. However, *Hs* can be predicted as a function of *Ht* with high ($R^2 = 0.70$ for endosperm type) and moderate ($R^2 = 0.57$ for heterotic group and 0.52 for inbreeding level) certainty. Finally, a significant correlation between *Ht* and *Hs* for kernel color was found for heterotic group ($r = 0.83$, $p = 0.04$) but not for inbreeding stage ($r = 0.41$, $p = 0.13$) or endosperm type ($r = 0.14$, $p = 0.80$).

A diverse set of regression coefficients (*b* values) differentiated between kernel traits based on the main source of variation used in building the respective regression models. Significant *b* values for biochemical traits ranged from a minimum of 0.53 ($p = 0.009$), when based on inbreeding stage, to a maximum of 0.99 ($p = 0.06$), when based on heterotic groups. The exception was the regression coefficient for biochemical traits based on endosperm type ($b = 1.4$).

The relationships between *Ht* and *Gst* were predominantly linear and mostly positive but not always significant (Fig. 5). None of the relationships between *Ht* and *Gst* were significant when heterotic group was used as a main source of variation, except in case of kernel physical traits ($r = 0.61$, $p = 0.007$). Population differentiation of kernel physical, biochemical, and nutrient traits but not color traits increased linearly and significantly as *Ht* increased when level of inbreeding was used as main source of variation. In this case, *Gst* can be

Table 3. Percent variance and its level of significance (*P*) due to main effects of five sources of variation [i.e., populations, genotypes(populations), endosperm type, heterotic group, and selfing stage], least squares mean, and coefficient of variation (CV%) for kernel physical traits of maize populations.

			Kernel physical traits								
Variance component	Category	Statistic	Weight	Density	Area	Perimeter	Width	Length	Major axis	Minor axis	Circularity
			mg	g cm^{-3}	mm^2	mm	mm	mm	mm	mm	%
Population		%variance	36.4	10.8	35.4	28.3	25.6	27.4	32.3	29.2	17.1
		P	0.01	0.03	0.01	0.02	0.02	0.02	0.02	0.01	0.04
		Mean	0.269	1.241	67.12	35.5	9.87	9.43	10.59	8.0	0.675
		CV%	18.5	4.1	16.4	10.4	9.7	9.5	10.6	9.15	10.24
Genotypes(populations)		%variance	4.1	27.2	2.8	3.3	2.2	2.1	1.4	3.4	1.3
		P	0.19	0.03	0.23	0.18	0.26	0.25	0.35	0.28	0.32
Endosperm type		%variance	5.9	33.2	1.9	1.6	2.1	2.3	1.9	1.2	1.1
		P	0.05	0.0001	0.07	0.25	0.18	0.12	0.23	0.43	0.98
	Opaque	Mean	0.25	1.2	63.01	32.75	9.5	9.1	10.32	7.81	0.68*
		CV%	17.9	3.4	16.1	10.5	9.5	9.7	10.6	8.9	10.6
	Translucent	Mean	0.27*	1.3*	68.31*	36.96*	9.9*	9.7*	10.71*	8.11*	0.64
		CV%	17.7	3.3	16.3	10.1	9.7	9.3	10.4	9.2	9.9
Heterotic group		%variance	3.9	2.5	2.2	2.3	2.4	2.5	1.8	1.2	1.2
		P	0.12	0.21	0.17	0.11	0.17	0.15	0.24	0.32	0.22
	Stiff stalk	Mean	0.275*	1.246	68.32	35.92	10.02	9.46	9.97	10.61	8.16
		CV%	21.2	4.0	17.1	9.3	9.8	9.3	10.4	9.8	9.8
	Non-stiff stalk	Mean	0.266	1.237	66.36	35.28	9.81	9.40	9.82	10.59	7.91
		CV%	16.2	4.0	15.8	11.0	9.5	9.6	10.7	8.5	10.5
Selfing		%variance	29.4	10.5	24.8	23.3	15.2	18.5	18.8	22.3	13.3
		P	0.005	0.001	0.007	0.006	0.006	0.005	0.007	0.001	0.001
	Open pollinated	Mean	0.306a†	1.22c	77.56a	38.44a	10.73a	10.11a	11.34a	8.71a	0.67b
		CV%	13.9	3.4	9.1	6.3	6.0	6.1	6.7	6.9	10.1
	S1‡	Mean	0.263b	1.24b	66.52b	35.81b	9.86b	9.43b	10.64b	7.92c	0.66b
		CV%	17.4	4.1	15.7	10.4	9.4	9.5	10.5	8.4	10.9
	S2	Mean	0.253b	1.22c	62.41c	33.07d	9.48b	9.04c	10.13c	7.82c	0.72a
		CV%	16.2	4.0	14.9	9.7	9.7	8.7	10.3	8.1	8.0
	S3	Mean	0.259b	1.24b	63.48c	34.21c	9.54b	9.13c	10.10c	7.97c	0.68b
		CV%	16.9	3.6	14.5	8.1	7.8	8.2	8.2	9.8	8.4
	S4	Mean	0.329a	1.28a	77.95a	37.71a	10.61a	10.05a	11.30a	8.76a	0.69b
		CV%	13.4	3.3	13.4	8.2	8.9	8.4	9.8	7.5	6.9
Total variance			79.7	84.2	67.1	58.8	47.5	52.8	56.2	57.3	23.0

*Significant at the 0.05 probability level. This indicates a significant difference between means within traits of opaque and translucent endosperm types or between stiff stalk and non-stiff stalk heterotic groups.

†Means within traits followed by the same letter do not differ significantly (Duncan's multiple range test, 5%).

‡S1 to S4 indicate the number of selfed generations.

predicted with decreasing certainty (R^2 = 0.41, 0.22, 0.12, and 0.1) for kernel physical, nutrient, color, and biochemical traits, respectively. Finally, positive and significant relationships between *Ht* and *Gst* were found for kernel physical (r = 0.48, p = 0.04) and color (r = 0.87, p = 0.02) traits, a positive and nonsignificant relationship (r = 0.17, p = 0.45) was found for kernel nutrients, and a negative and significant relationship (r = −0.58, p = 0.02) was found for kernel biochemical traits when endosperm type was used as a main source of variation.

A preliminary comparison between regression models for kernel traits when different sources of variation were used in developing these models suggested that higher levels of population differentiation of kernel physical traits are expected as *Ht* increases when based on heterotic group and level of inbreeding (b = 1.2 in both cases) than when based on endosperm type (b = 0.6). Similarly, larger *Gst* for biochemical and nutrient traits is expected as *Ht* increases, with b = 0.54 and 0.42, respectively, based on level of inbreeding as a main source of variation. The negative and significant correlation between *Ht* and *Gst* for biochemical traits when based on endosperm type suggested that *Gst* is expected to decline (b = −0.79) as *Ht* increases while *Gst* for kernel color is expected to increase by more than one unit (b = 1.1) with each unit increase in *Ht*.

All correlations between *Hs* and *Gst* were negative and significant (Fig. 6), with two exceptions. Two nonsignificant correlations were found for kernel color when heterotic group (r = −0.29) and endosperm type (r = −0.36) were used as sources of variation. A wide range of certainties (R^2 ranged from 0.6 to 0.83) in predicting *Gst* as a function of diversity within populations can be deduced from the regression models. Population differentiation can be predicted with large and significant certainties for kernel physical (R^2 = 0.60) and biochemical traits (R^2 = 0.83) but with relatively smaller and less significant certainties for nutrients (R^2 = 0.28) when heterotic group was used as a main source of variation.

Table 4. Percent variance and its level of significance (*P*) due to main effects of five sources of variation [i.e., populations, genotypes(populations), endosperm type, heterotic group, and selfing stage], least squares mean, and coefficient of variation (CV%) for kernel biochemical traits of maize populations.

			Biochemical traits							
Variance component	Category	Statistic	Moisture %	Protein %	Lys %	Met %	Cys %	Oil %	Starch %	C:N ratio
Populations		%variance	17.2	12.8	5.2	8.2	14.4	13.8	11.1	12.5
		P	0.01	0.02	0.04	0.03	0.02	0.02	0.02	0.02
		Mean	9.33	12.92	0.415	0.293	0.228	4.93	67.29	25.09
		CV%	26.5	15.87	9.46	14.41	10.77	12.56	2.89	17.93
Genotypes(populations)		%variance	22.5	11.3	33.8	21.3	9.9	21.9	21.5	10.4
		P	0.08	0.14	0.08	0.12	0.13	0.11	0.11	0.13
Endosperm type		%variance	31.9	5.2	16.8	7.5	9.5	23.8	10.4	5.2
		P	0.005	0.12	0.01	0.05	0.06	0.006	0.02	0.05
	Opaque	Mean	7.5	12.9	0.44*†	0.31*	0.21	5.3*	66.5	25.6*
		CV%	21.7	15.9	8.2	13.5	11.2	12.5	2.7	17.9
	Translucent	Mean	10.4*	12.9	0.40	0.28	0.24*	4.8	67.6*	24.5
		CV%	21.5	15.8	9.8	14.8	10.4	11.2	2.9	17.8
Heterotic group		%variance	2.9	2.6	3.2	2.8	1.9	1.5	1.1	0.98
		P	0.25	0.48	0.27	0.25	0.35	0.35	0.42	0.45
	Stiff stalk	Mean	9.8 *	12.9	0.40	0.29	0.23	4.9	67.4	25.1
		CV%	31.8	16.2	9.2	014.5	11.3	13.8	2.9	18.8
	Non-stiff stalk	Mean	9.0	12.9	0.42*	0.29	0.23	4.9	67.2	25.1
		CV%	21.1	15.7	9.6	14.3	10.5	11.7	2.9	17.4
Selfing		%variance	15.7	26.8	12.1	22.5	27.8	18.4	19.4	28.8
		P	0.001	0.05	0.01	0.009	0.006	0.005	0.009	0.005
	Open pollinated	Mean	8.6c†	11.5d	0.42a	0.27c	0.21b	6.1a	67.4b	28.3a
		CV%	19.7	14.4	7.3	11.9	9.6	10.2	2.5	14.1
	S1‡	Mean	9.2b	13.1b	0.42a	0.32a	0.23b	4.9b	67.1c	24.8b
		CV%	24.1	15.4	9.8	14.4	10.1	11.6	3.0	16.6
	S2	Mean	8.4c	12.7b	0.42a	0.29b	0.23b	4.9b	67.3b	26.0ab
		CV%	19.5	18.1	9.5	15.9	12.1	13.0	3.1	20.4
	S3	Mean	8.3c	12.1c	0.40b	0.27c	0.22b	4.9b	67.9a	27.0a
		CV%	18.7	16.9	8.3	14.3	11.6	11.5	2.6	19.1
	S4	Mean	12.9a	13.9a	0.42a	0.32a	0.25a	4.6c	67.1c	22.1c
		CV%	19.9	10.2	8.4	10.8	6.9	11.9	2.3	11.8
Total variance			90.2	58.7	71.1	62.3	63.5	79.4	63.5	57.9

*Significant at the 0.05 probability level. This indicates a significant difference between means within traits of opaque and translucent endosperm types or between stiff stalk and non-stiff stalk heterotic groups.

†Means within traits followed by the same letter do not differ significantly (Duncan's multiple range test, 5%).

‡S1 to S4 indicate the number of selfed generations.

The level of inbreeding was a strong source of variation to use in predicting *Gst* as a function of *Hs* for all kernel traits except nutrients. All correlation coefficients between *Gst* and *Hs* were negative and significant although they ranged in magnitude from large ($r = -0.74$ for kernel physical traits) to small ($r = -0.25$ for nutrients); consequently, the R^2 values were medium (0.51 to 0.55) to very small (0.06). Correlations between *Gst* and *Hs* were significant for kernel physical traits ($r = -0.59$) and biochemical traits ($r = -0.91$), marginally significant ($r = -0.40$) for nutrients, and nonsignificant ($r = -0.36$) for color traits when endosperm type was used as a main source of variation. Therefore, *Gst* can be predicted with reasonable certainty for biochemical traits ($R^2 = 0.82$) and to some extent for kernel physical traits ($R^2 = 0.35$).

The expected rate of decline in *Gst* (expressed as *b*) as a function of *Hs* depended on the trait in question and on whether heterotic group, level of inbreeding, or endosperm type was used as a main source of variation for its prediction. The expected rate of decline in *Gst* for kernel physical traits is small ($b = -0.84$) when based on endosperm type and large ($b = -1.34$ and -1.35) when based on inbreeding level and heterotic group. A similar trend was observed for biochemical traits while different trends were observed for the nutrients and kernel color traits.

DISCUSSION

The commercial germplasm base of maize is not considered to be heterogeneous (Singh et al., 2011) nor does it represent unimproved or unselected germplasm. The latter is considered to be an important genetic resource needed to protect the gains in maize yield, drive increases in genetic yield potential (Ortiz et al., 2010), and improve the nutritional quality of the crop (Mittelmann et al., 2011). End-use quality (Ortiz-Monasterio et al., 2007) as well as early performance and growth of agricultural crops are

Table 5. Percent variance and its level of significance (*P*) due to main effects of five sources of variation [i.e., populations, genotypes(populations), endosperm type, heterotic group, and selfing stage], least squares mean, and coefficient of variation (CV%) for kernel macro- and micronutrients in maize populations.

Variance component	Category	Statistic	Macro- and micronutrients										
			C	N	Ca	Cu	Fe	K	Mg	Mn	P	S	Zn
			%	%	Mean, g kg^{-1}								
Population		%variance	12.5	12.5	8.5	8.7	14.9	31.5	21.7	17.9	23.2	12.1	14.2
		P	0.01	0.01	0.02	0.01	0.01	0.02	0.02	0.02	0.01	0.01	0.03
		Mean	45.5	1.87	0.45	0.0016	0.0236	3.62	1.39	0.067	3.45	1.24	0.0238
		CV%	0.72	16.43	40.4	82.3	27.66	14.16	10.8	20.74	10.7	12.4	16.8
Genotypes(populations)		%variance	26.2	10.9	3.9	0.85	1.2	7.3	11.3	11.2	10.9	15.9	4.8
		P	0.04	0.06	0.09	0.22	0.28	0.11	0.05	0.06	0.05	0.05	0.17
Endosperm type		%variance	4.2	2.3	2.5	2.5	1.8	12.1	2.9	1.7	3.6	2.1	2.5
		P	0.06	0.23	0.34	0.54	0.82	0.02	0.09	0.08	0.09	0.42	0.56
	Opaque	Mean	45.7*†	1.89*	0.46	0.0017	0.0235	3.79*	1.42*	0.069*	3.52*	1.24	0.0242
		CV%	0.77	16.8	40.9	82.2	29.2	13.1	10.9	21.1	10.7	12.9	17.6
	Translucent	Mean	45.2	1.82	0.45	1.4	2.37	3.53	1384	0.065	3.41	1.25	24.7
		CV%	0.68	16.1	40.1	81.9	26.8	14.1	10.6	20.4	10.5	12.1	16.3
Heterotic group		%variance	0.23	0.35	0.57	2.8	1.7	0.87	1.4	1.9	1.1	1.4	3.2
		P	0.37	0.28	0.24	0.25	0.26	0.37	0.26	0.25	0.27	0.26	0.24
	Stiff stalk	Mean	45.5	1.88	0.45	0.0017	0.0246	3.65	1.41	0.069	3.48	1.26	0.0245
		CV%	0.81	17.1	72.5	78.5	29.1	14.6	10.1	19.1	9.9	12.3	17.1
	Non-stiff stalk	Mean	45.5	1.86	0.45	0.0014	0.0230	3.61	1.38	0.066	343	1.23	0.0235
		CV%	0.66	16.1	41.3	84.0	26.3	13.8	11.2	21.6	11.1	12.5	16.4
Selfing		%variance	12.2	13.6	4.2	10.6	13.6	10.5	23.2	24.4	21.9	29.8	5.2
		P	0.06	0.01	0.29	0.05	0.05	0.03	0.008	0.007	0.009	0.006	0.05
	Open pollinated	Mean	46.0a†	1.65d	0.45	0.0014b	0.0271a	3.43c	1.34b	0.064c	3.29c	1.23b	0.0253a
		CV%	0.67	14.4	39.8	57.8	23.3	9.5	10.2	21.2	9.9	12.5	16.2
	S1‡	Mean	45.5b	1.87b	0.46	0.0014b	0.0230b	3.69a	1.41a	0.068a	3.49a	1.24b	0.0237b
		CV%	0.72	15.7	38.8	85.3	26.8	14.3	11.1	20.7	11.0	12.0	15.9
	S2	Mean	45.4b	1.81b	0.43	0.0012d	0.0231b	3.59b	1.36b	0.064c	3.38b	1.21b	0.0233b
		CV%	0.79	18.2	74.4	87.6	33.0	12.6	11.7	24.8	11.2	14.9	21.2
	S3	Mean	45.4b	1.74c	0.43	0.0023a	0.0260a	3.66a	1.37b	0.066b	3.37b	1.21b	0.0249a
		CV%	0.71	18.0	42.8	58.2	26.7	12.1	10.0	20.1	9.7	13.2	17.0
	S4	Mean	45.8a	2.07a	0.44	0.0013c	0.0240b	3.43c	1.40a	0.068a	3.45a	1.32a	0.0237b
		CV%	0.62	10.6	42.0	86.1	23.9	14.0	8.1	15.7	8.1	8.3	14.7
Total variance			55.3	39.7	19.7	25.5	33.2	62.3	60.5	57.1	60.7	61.3	29.9

*Significant at the 0.05 probability level. This indicates a significant difference between means within traits of opaque and translucent endosperm types or between stiff stalk and non-stiff stalk heterotic groups.

†Means within traits followed by the same letter do not differ significantly (Duncan's multiple range test, 5%).

‡S1 to S4 indicate the number of selfed generations.

affected by seed quality (Goggi et al., 2008). Therefore, understanding the relationships among and within nutritionally and industrially important traits in the maize kernel with the concomitant goal of improving its end-use value is increasingly an important objective for breeders, agronomists, and farmers. The objective of the breeding program that generated the germplasm was to develop high-quality maize varieties with enhanced levels of protein and the essential amino acids cysteine, lysine, and methionine without compromising the agronomic competitive ability of the newly developed varieties.

The germplasm evaluated in this study (Table 1) included three (B73, LH82, and Mo17) of the seven inbred lines considered as most important commercial maize genetic resources in the United States (Singh et al., 2011). Two of the inbred lines represented in the genetic background of this germplasm (B73 and Mo17) and their respective mutants provide excellent variation in endosperm texture and hardness as well as associated variation in zein composition (Fox and Manley, 2009). Exotic germplasm was represented in the genetic background of a large portion of the evaluated germplasm, including Corn Belt, Native American, and "tropical" Central and South American maize landraces. Although the integration of tropical maize germplasm into local breeding programs has been hindered by lack of adaptation of this germplasm to Corn Belt environments (Scott and Blanco, 2009), it was successfully incorporated into the genetic background of some populations in this germplasm. It is considered a potentially valuable source of genetic diversity for the improvement of U.S. Corn Belt varieties. Maize landraces have not been selected to perform under high inputs; however, they have not undergone inbreeding and are characterized by having more protein and less starch as compared with maize inbreds (Flint-Garcia et al., 2009). Many studies (Harrigan et al., 2010; Skogerson et al., 2010) suggested a high level of natural variability inherent to the biochemical composition of maize. Cultivation in different regions and under contrasting management practices may have contributed more than genetic modification to compositional differences between maize varieties (Harrigan et al., 2010).

The study of a large number of accessions (1348) and kernels (13,480) with different combinations of genetic backgrounds, heterotic groups, inbreeding levels, endosperm types, and nutritional quality represents the first extensive grain quality characterization performed on this diverse germplasm. This wide genetic base and the large number of kernel traits (31) in the study provided an opportunity to identify sources of variation, analyze genetic diversity, and quantify multivariate relationships between several physical, biochemical, nutrient, and color traits of the maize kernel.

Table 6. Percent variance and its level of significance (*P*) due to main effects of five sources of variation [i.e., populations, genotypes(populations), endosperm type, heterotic group, and selfing stage], least squares mean, and coefficient of variation (CV%) for kernel color descriptors of maize populations.

Variance component	Category	Statistic	Kernel color descriptor† *L**	*a**	*b**
Population		%variance	25.9	19.7	18.4
		P	0.01	0.01	0.01
		Mean	82.9	–3.7	9.2
		CV%	3.2	16.07	10.9
Genotypes(populations)		%variance	5.5	10.5	5.4
		P	0.12	0.06	0.14
Endosperm type		%variance	25.2	39.5	46.5
		P	0.04	0.02	0.03
	Opaque	Mean	81.6	–3.4	9.1
		CV%	3.2	14.2	10.2
	Translucent	Mean	83.5*†	–3.9*	9.5*
		CV%	3.1	15.9	8.5
Heterotic group		%variance	5.6	5.3	2.8
		P	0.25	0.24	0.31
	Stiff stalk	Mean	81.5	–3.5	9.3
		CV%	3.4	17.1	9.3
	Non-stiff stalk	Mean	83.2	–3.9*	9.3
		CV%	3.1	15.2	8.9
Selfing		%variance	32.1	7.4	17.6
		P	0.08	0.09	0.09
	Open pollinated	Mean	80.1b‡	–3.6a	9.9a
		CV%	3.4	15.7	9.2
	S1§	Mean	82.9a	–3.7a	9.2b
		CV%	3.1	16.2	8.8
	S2	Mean	84.6a	–3.9b	8.7c
		CV%	2.6	12.3	8.4
	S3	Mean	83.2a	–3.6a	9.1b
		CV%	3.1	18.3	10.5
	S4	Mean	80.4b	–3.5a	9.4b
		CV%	2.9	14.4	7.2
Total variance			94.3	82.4	90.7

*Significant at the 0.05 probability level. This indicates a significant difference between means within traits of opaque and translucent endosperm types or between stiff stalk and non-stiff stalk heterotic groups.

†*L**, darkness–lightness; *a**, redness–greenness; *b**, yellowness–blueness.

‡Means within traits followed by the same letter do not differ significantly (Duncan's multiple range test, 5%).

§S1 to S4 indicate the number of selfed generations.

Multivariate and Near-Infrared Spectroscopy Analyses

The comprehensive characterization of this germplasm was achieved by combining rapid testing for grain physical, biochemical, nutrient, and color quality traits using a number of analytical and multivariate statistical analysis procedures. Computerized seed imagery and several multivariate calibration and validation statistical techniques, such as ANN, PCA, PLS, and variance components estimation, are instrumental in extracting the desired biological, chemical, and physical information from large datasets (Payne et al., 2007). All of these techniques have been successfully used in building and analyzing the database in the current study and in testing the reliability of its results. However, regardless of the multivariate statistical analysis procedures used, there appears to be a complex interaction between the expression of these traits that depends on the genetic background (Cañas et al., 2011).

The NIRS predictions approached the accuracy of reference analytical tests we performed on subsamples of the germplasm collection for protein ($r = 0.97$, $p < 0.01$), oil ($r = 0.97$, $p < 0.01$), and starch ($r = 0.93$, $p < 0.03$) in this study. Similar accuracy levels have been reported on quality screening of early-generation material in cereal breeding programs (Osborne, 2001; Spielbauer et al., 2009), protein, tryptophan, and lysine in maize (Rosales et al., 2011), protein and starch in common bean (*Phaseolus vulgaris* L.) (Hacisalihoglu et al., 2010), and protein and several amino acids in peanut (*Arachis hypogaea* L.) (Wang et al., 2012).

Kernel Physical Traits

The PCAs (Fig. 2; Table 2) allowed the identification of single traits or groups of traits linked, directly or indirectly, to other traits or to populations in a multivariate and complex manner (Cañas et al., 2011). The physical shape and structure of kernels have been shown to affect hardness (Fox and Manley, 2009). Dent maize is softer than flint although there could be soft and hard endosperm within a single kernel. Kernels with opaque endosperm had smaller dimensions and were more circular with lower moisture and higher protein contents than those with translucent endosperm, despite the similarity in mean protein contents of opaque and translucent endosperm (12.9) (Table 4), the opaque endosperm was associated with larger protein content on PC1 (Fig. 2), which explained 0.47 of total variation.. Protein content (mean 12.92%) but not starch content (67.29%) plays a major role in determining density (and consequently hardness) of maize kernels although it makes up a much smaller content than starch (Fox and Manley, 2009).

Kernel weight and density did not display strong association with the level of inbreeding or selection (Fig. 2) as suggested by Scott et al. (2008). Evidently, the relationships between these and other kernel physical traits are population dependent (Seebauer et al., 2010). Nevertheless, physical differences in kernel density, hardness, and resistance to grinding as well as the attending differences in chemical composition may affect livestock performance (Prandini et al., 2011) whereas kernel hardness traits are negatively correlated with feed conversion and ruminal propionate concentration in livestock (Moore et al., 2008).

Kernel Biochemical and Nutrients Composition

Several authors (Nuss and Tanumihardjo, 2010) reported on "typical" kernel composition values for the commodity yellow dent maize (approximately 72.0% starch, 9.5% protein, 4.5% oil, 1.5% ash, and 2.5% sugars on dry matter basis); the average starch, protein, and oil estimated for this germplasm (67.29, 12.92, and 4.93%, respectively; see Table 4) are above most mean values reported in the literature (Harrigan et al., 2010). However, kernel moisture content, which is one of a few traits significantly

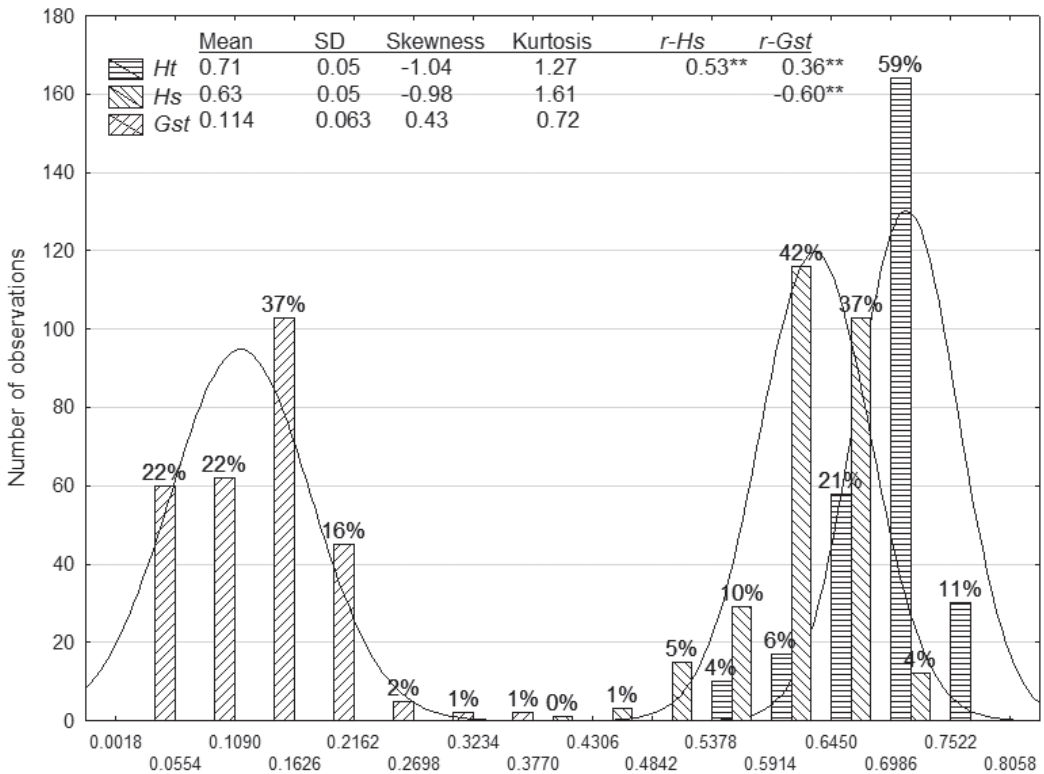

Figure 3. Frequency distribution of phenotypic total diversity (*Ht*) and its components (diversity within populations [*Hs*] and the population differentiation [*Gst*] coefficient) based on 31 physical, biochemical, chemical, and color traits of 13 maize populations. **Significant at the 0.01 probability level. *r-Hs*, Correlation coefficient between *Hs* and *Ht*; *r-Gst*, correlation coefficient between *Gst* and each of *Ht* and *Hs*.

influenced by the environment at harvest (mean 9.33%), displayed the largest variation (CV = 26.5%).

Modern maize breeding and selection targeted seed size and starch content, which became of interest as a source of ethanol (Fox and Manley, 2009) at the expense of protein content and its associated amino acids (Flint-Garcia et al., 2009). Variation in biochemical composition (CV ranged from 2.89% for starch to 15.87% for protein; see Table 4) provides an opportunity for selecting accessions with improved kernel quality, including higher contents of the essential amino acids cysteine, lysine, and methionine. Two inbreds (B73 and Mo17) included in this study typically have low protein (11.06 and 11.9%, respectively) and large starch (73.31 and 71.94%, respectively) contents (Flint-Garcia et al., 2009). The population derived from B73 loaded on the positive side of PC1 (Fig. 2) with larger protein (>12.92%) and smaller starch (<67.29%) contents than the overall mean whereas the population derived from Mo17 loaded on the negative side of PC1 with larger protein and smaller starch contents than the respective means.

This germplasm may provide an opportunity to simultaneously select accessions with high starch and oil contents; both biochemicals had positive and large loadings on PC1 and PC2, respectively (Table 2), and are localized in different parts of the kernel (Chander et al., 2008). Although C and N assimilations are strongly linked, a larger C:N ratio (Fig. 2), which represents the capacity of the kernel to store more carbohydrates, may be associated with the accumulation of starch over oil (Cañas et al., 2011). This can be deduced from loadings of starch and oil on and correlation coefficients of C, N, and C:N with PC1 and PC2 (Table 2). Maize genotypes capable of maintaining a higher C:N of assimilate transport can produce kernels with higher starch:protein ratios such as the four populations in quadrant II (Fig. 2) that were mainly in the stiff stalk heterotic group, (Seebauer et al., 2010).

Average amino acid contents (Table 4) were comparable to or higher than those reported for a high-yielding hybrid dent, hybrid semiflint, quality protein maize variety, and unimproved maize landrace (Fufa et al., 2003). Dynamics of cysteine and methionine are closely correlated with each other and with protein (Fig. 1); lysine, on the other hand, is slightly different and all three amino acids can be estimated from reference protein values (Rosales et al., 2011) because

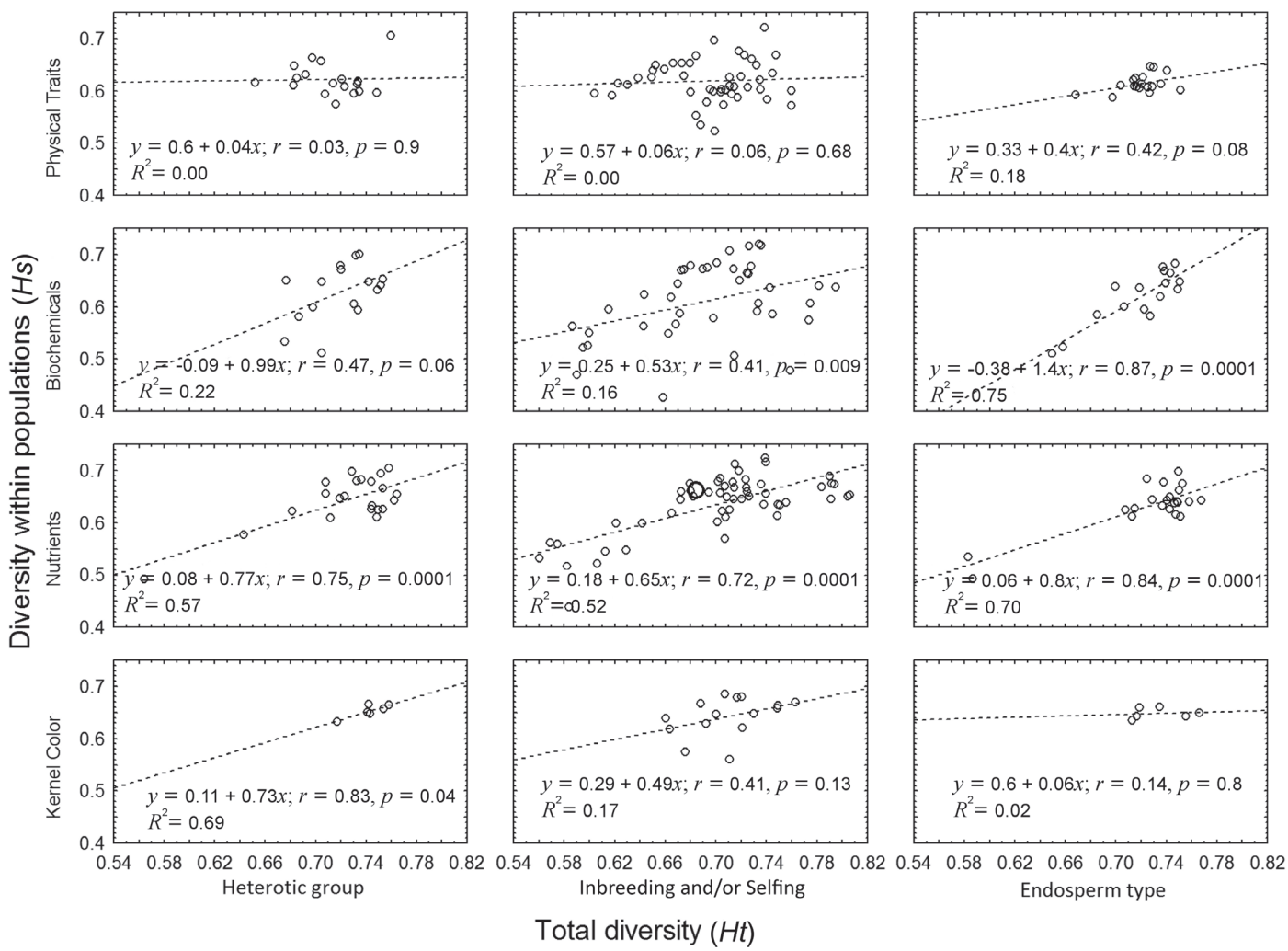

Figure 4. Statistical relationships and test of significance of total diversity (*Ht*) and diversity within populations (*Hs*) for each of kernel physical, biochemical, chemical, and color traits based on heterotic group, inbreeding status, and endosperm type in 13 maize populations.

of the existence of strong positive correlations as can be deduced from PCA (Table 2) and trait associations (Fig. 2).

The large loadings of macro- and micronutrients on PC1 (except Ca) and their significant correlation coefficients with PC1 derived from biochemical traits (Table 2 and as verified in Fig. 2) indicated co-localization and strong association with protein, especially in kernels with opaque endosperm. Macronutrients were less variable than micronutrients, with C being the least (CV = 0.72%), N (CV = 16.43%) intermediate, and Ca (CV = 40.4%) the most variable (Table 5). Most if not all micronutrient concentrations in the maize kernel (Ortiz-Monasterio et al., 2007) and other field crops (Murphy et al., 2008) are limited by their availability in the soil. Copper displayed the largest variation (CV = 82.3%) while S, a critical micronutrient in the biosynthesis of protein, cysteine, and methionine (Bellaloui et al., 2011), is more conserved and displayed the smallest variation (CV = 12.4%).

Kernel Color Patterns

Modern maize varieties show little variation for kernel color (Floyd et al., 1995; Orak et al., 2010). Kernel color patterns of this germplasm, as quantified by the $L^\star$, $a^\star$, and $b^\star$ color space descriptors, were more variable than commercial yellow dent maize (Orak et al., 2010). At the whole germplasm level (Table 6), breeding lines in this germplasm exhibited large mean values but small variation for $L^\star$ (light-colored kernels; CV = 3.2%), large negative mean values and large variation for $a^\star$ (red; CV = 16.07%), and large mean values and moderate variation for $b^\star$ (yellow; CV = 10.9%).

Loadings and associations of $L^\star$, $a^\star$, and $b^\star$ on PC1 and PC2 (Fig. 2) are indicative of the tremendous color variability in this germplasm. Typically, $L^\star$ and (more negative) $a^\star$ are positively correlated while $L^\star$ and $b^\star$ are negatively correlated (Itle and Kabelka, 2009). There were significant negative ($r = -0.69$) and positive ($r = 0.63$) correlation coefficients of $L^\star$ and $b^\star$ with PC1 derived from physical kernel traits (Table 2). Loadings on PC1 and PC2 indicated a strong positive and a weak negative association between $L^\star$ and $b^\star$, respectively. However, the negative association between $L^\star$ and $a^\star$ on both PCs is a result of the slightly negative $a^\star$ values (mean of −3.7). Higher values of red color (negative $a^\star$) (Fig. 2) and negative correlation ($r = -0.31$) with PC2 derived from

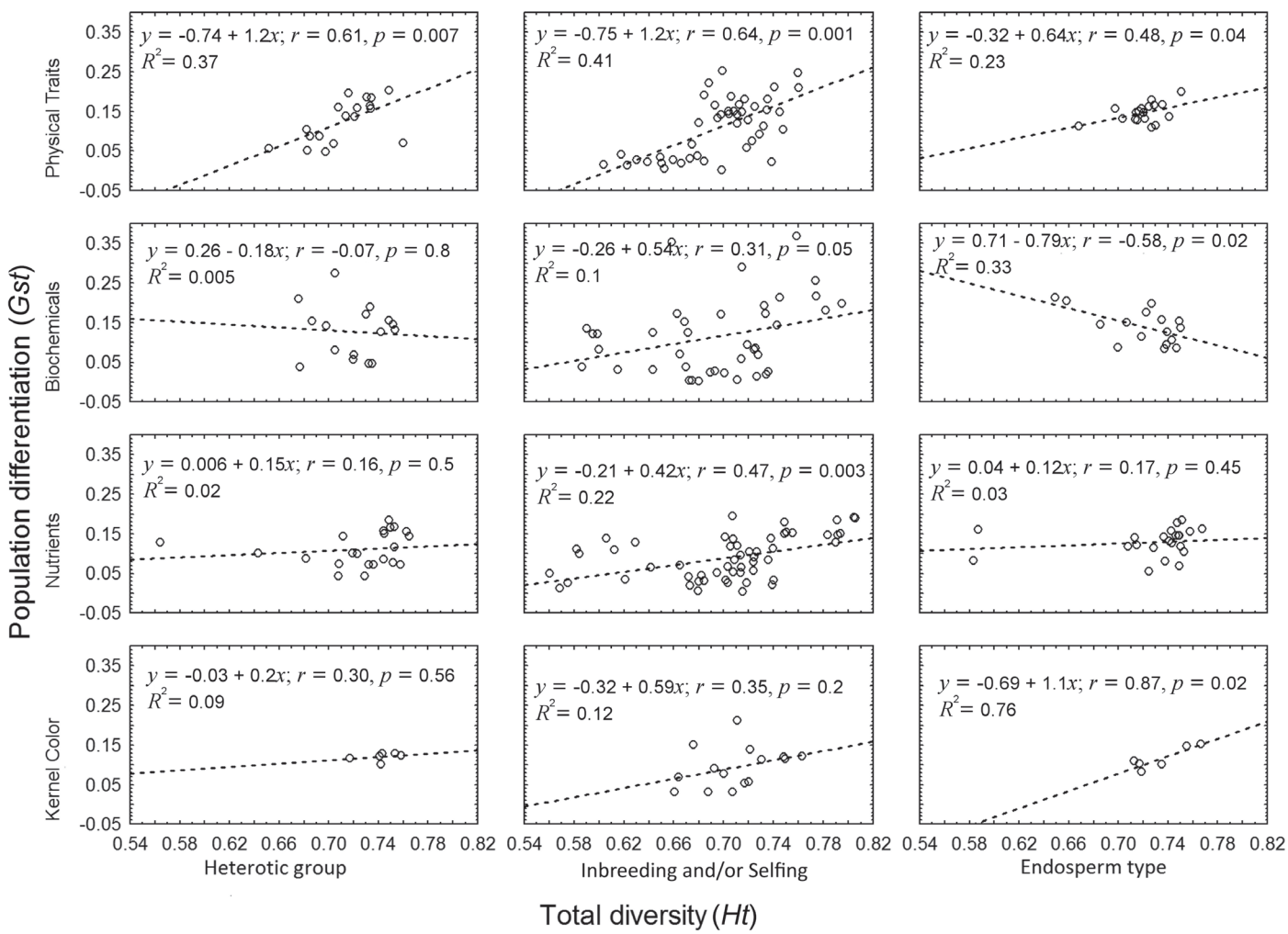

Figure 5. Statistical relationships and test of significance of total diversity (*Ht*) and population differentiation (*Gst*) coefficient for each of kernel physical, biochemical, chemical, and color traits based on heterotic group, inbreeding status, and endosperm type in 13 maize populations.

biochemical traits with large loading of oil (0.87) (Table 2) may indicate higher content and activity of peroxidase and lipoxygenase; these enzymes are positively correlated with oil, protein, and Fe contents (Orak et al., 2010). Also, increase in yellow color ($b^{\star}$) may reflect higher carotenoid content or lower C18:2 and C18:3 fatty acid contents (Orak et al., 2010; Kuhnen et al., 2011) and is positively correlated on PC1 with starch content (Fig. 2).

Physicochemical Trait Relationships

Carbon and N assimilations are strongly linked (Cañas et al., 2011) and are associated with some physical traits (e.g., kernel weight) and with most biochemical and all nutrients, except Ca (Table 2). The correlation between protein content and density, although expected to be positive (Seebauer et al., 2010), is population dependent and is influenced by the endosperm type. Denser kernels (e.g., population NG-Mo508-Mo506) are expected to have higher protein (Flint-Garcia et al., 2009) and hence the negative correlation between protein content and kernel weight on PC1 (Fig. 2).

Interrelationships between kernel weight, starch, protein, and oil in this germplasm are complex (Fig. 2; Table 2) and, for the most part, conform to those frequently reported in the literature (Chander et al., 2008). However, population-dependent positive loadings of oil and starch on PC1 may suggest the likelihood of simultaneous selection for both traits in some populations. The interrelationships between cysteine, lysine, and methionine and their interaction with protein, oil, starch, and nutrients are in agreement with earlier reports on maize (Olsen et al., 2003; Jamal and Abdin, 2010) and soybean [*Glycine max* (L.) Merr.] (Bellaloui et al., 2011). At a multivariate level, cysteine, lysine, and methionine were correlated with PC1 derived from nutrients, which has (in decreasing order) high loadings of Mn, Mg, P, S, Zn, K, Fe, and Cu (Table 2). These correlations may suggest a possible role of these nutrients in seed biochemical composition (Bellaloui et al., 2011). The strong associations of Fe and Zn with cysteine, lysine, and methionine and with protein and the strong association of S with cysteine and methionine (Table 2; Fig. 2) reinforce the importance of macro- and

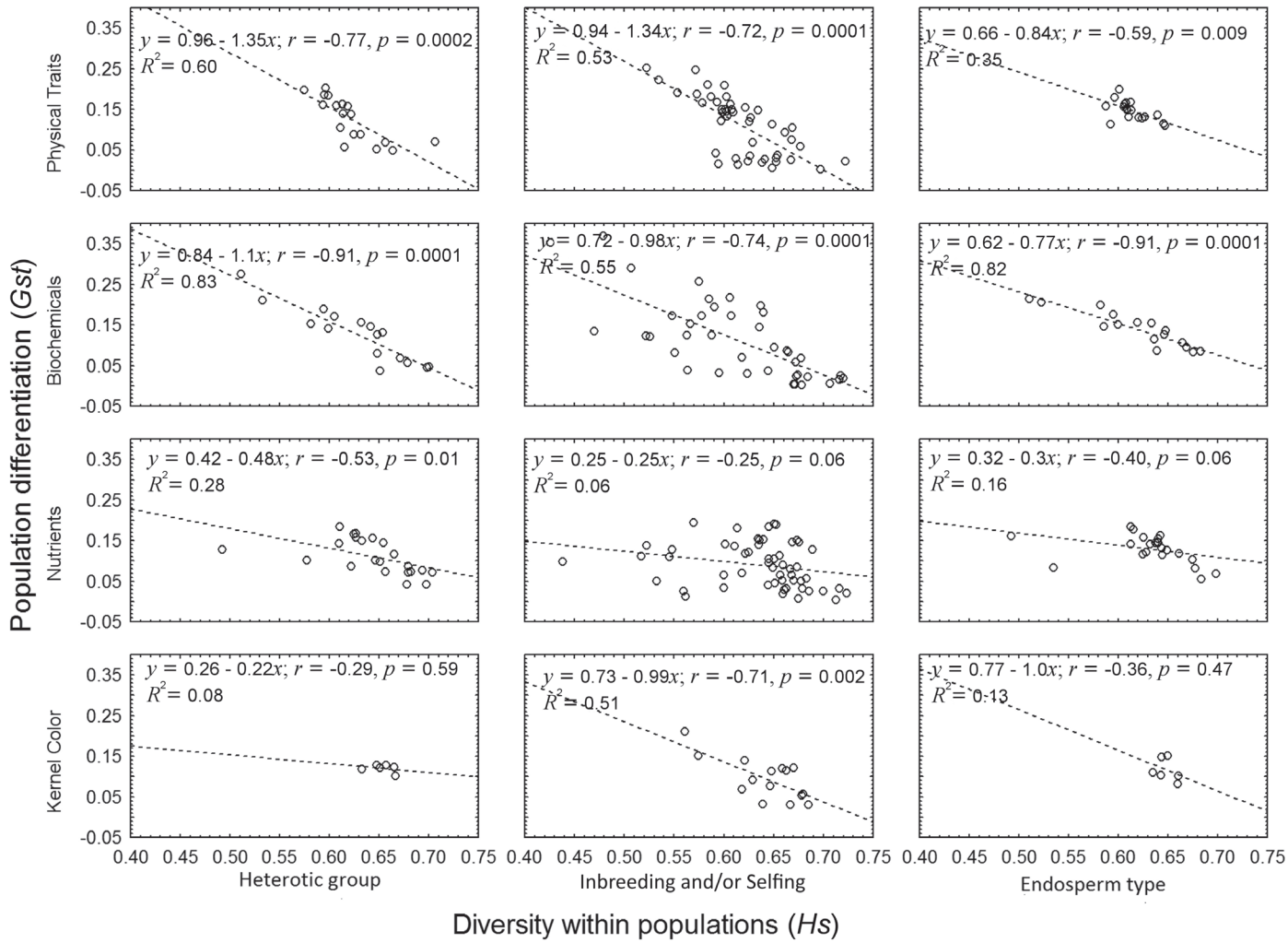

Figure 6. Statistical relationships and test of significance of diversity within populations (*Hs*) and population differentiation (*Gst*) coefficient for each of kernel physical, biochemical, chemical, and color traits based on heterotic group, inbreeding status, and endosperm type in 13 maize populations.

micronutrients, not only for their specific nutritional value (Welch and Graham, 2004) but also as determinants of protein quality (Jamal and Abdin, 2010) and for their direct or indirect role in oil biosynthesis (Thomison et al., 2003; Bellaloui et al., 2011).

Sources of Trait Variation

Historically, the composition of maize kernel exhibited clear trends with time (Scott and Blanco, 2009), with protein decreasing and starch increasing. Similarly, on a per tissue mass basis, the level of the amino acids lysine and methionine and nutrients also decreased over time (Welch and Graham, 2004; Rosales et al., 2011). Food, feed, and industrial applications require maize kernels with certain levels of desirable physical, biochemical, nutrient, and color traits (Seebauer et al., 2010; Taboada-Gaytan et al., 2010). Some of these traits, such as amino acids, need in-depth analysis to identify their sources and quantify their levels of variation; other traits, such as starch, oil, and crude protein (Kaur et al., 2010), can be assessed visually and indirectly using covariance analysis. In this study, the multivariate analysis procedures were instrumental in identifying sources of variation and in partitioning total variation into its components to facilitate the selection and potentially combining desired levels of quality traits appropriate for a particular end-use application (Nuss and Tanumihardjo, 2010).

On average, the main sources of variation in this study explained 60, 68, 46, and 89% of the total variation in physical, biochemical, nutrient, and kernel color traits, respectively. The careful choice of parents and the crossing and inbreeding scheme succeeded in generating large levels of variation among genotypes within populations for most traits (e.g., 27.2% for kernel density, 33.8% for lysine, 15.9% for S). However, variation among populations was predominantly the largest for kernel weight (36.4%), Fe (14.9%), and Zn (14.2%).

Quality Assessment

An important concept in improving kernel quality of maize is increasing the contents of cysteine, lysine, and methionine as limiting essential amino acids through breeding and selection (Flint-Garcia et al., 2009) to replace expensive supplements in maize-based human food and animal feed (Scott and Blanco, 2009). However, kernel contents of protein, oil, and starch as well as nutrient concentrations and kernel physical characteristics are important quality attributes (Moore et al., 2008; Goggi et al., 2008). The simultaneous improvement of some of the biochemical traits in maize while maintaining high grain yield, although difficult, is considered feasible (Mittelmann et al., 2011). There is ample variation for and interrelationships among multiple traits in this germplasm that can be manipulated by breeding and selection to develop maize varieties with a desired set of trait combinations. This analysis indicated the feasibility of combining high protein and increased essential amino acid contents, high starch and high oil contents, or high macro- and micronutrient concentrations while at the same time maintaining acceptable kernel weight and incorporating desired kernel color attributes in new varieties (Chander et al., 2008; Scott et al., 2008). High levels of lysine and methionine were achieved in some of the opaque breeding lines (Fig. 2; Table 4). Even with such breakthroughs in high quality varietal development (Harrigan et al., 2010), the overall maize nutritional quality may not reach the level of wheat (*Triticum aestivum* L.) (Murphy et al., 2008) but may become slightly better than that of rice (*Oryza sativa* L.) as human food (Nuss and Tanumihardjo, 2010).

Germplasm accessions derived from landraces (Orak et al., 2010) are a rich genetic resource for high pigment levels. Incorporating genes for kernel color in new commercial varieties may improve food, feed, and industrial quality of maize products. In addition, these "color markers" exhibit significant relationships with oil, crude fiber, and protein in the kernel (Kaur et al., 2010) and may serve as markers in breeding and selecting high-quality maize varieties. Colored maize kernels, as quantified by $L^\star$, $a^\star$, and $b^\star$ (Table 6), can be used for the development of grain-based functional foods or as natural colorants besides their value as genetic resources in breeding programs (Kuhnen et al., 2011).

CONCLUSION

One of the most serious unintended consequences of long-term selection in the context of commercial maize breeding was the loss of genetic variation due to excessive gametic linkage disequilibrium. To reverse this trend, maize breeders will need readily available and widely variable germplasm to continue yield increases and to improve multiple quality traits. We identified sources of variation of single and multiple quality traits, determined the level of genetic diversity present in a wide pool of maize germplasm, and quantified phenotypic variation that may be useful for developing high-quality maize varieties. The germplasm and information are also of value for genetic studies to investigate the genetic architecture of relevant quality traits. We developed a relational database and constructed a complex array of direct and indirect relationships between and within physical, biochemical, nutrient, and color traits of maize kernels that can be used for further research and for in-depth understanding of its diversity structure. We found genotype- and population-specific differences for multiple traits, thus making it applicable to focus on a set of traits that are correlatively expressed for breeding and selection purposes.

Acknowledgments

We thank Jana Rinke for her skilled laboratory work and Jon Starr for preparing the figures. This research was funded by The Department of Agriculture, Trade, and Consumer Protection, State of Wisconsin, USDA Agricultural Marketing Resource Center, Ames, IA, and by USDA-ARS Project No. 3645-61600-001-00D, Morris, MN. We acknowledge the services and collaboration of The Grain Quality Testing Laboratory at Iowa

State University and The Agricultural Research Station Chemical Laboratory, University of Missouri-Columbus, MO. The breeding program was managed by WG during his tenure at Michael Fields Agricultural Institute, East Troy, WI, with funding from USDA-ARS and in collaboration with Dr. Linda Pollak, Corn Geneticist, Corn Insect and Crop Genetics Unit, Ames, IA (Retired). The use of trade, firm, or corporation names in this publication is for information and convenience of the reader. Such use does not constitute an official endorsement or approval by the United States Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable. The USDA is an equal provider and employer.

References

Bellaloui, N., J.R. Smith, A.M. Gillen, and J.D. Ray. 2011. Effects of maturity, genotypic background, and temperature on seed mineral composition in near-isogenic soybean lines in the early soybean production system. Crop Sci. 51:1161–1171. doi:10.2135/cropsci2010.04.0187

CAMO. 2011. The Unscrambler software, release 10.1. CAMO Software, Oslo, Norway.

Cañas, R.A., N. Amiour, I. Quillere, and B. Hirel. 2011. An integrated statistical analysis of the genetic variability of nitrogen metabolism in the ear of three maize inbred lines (*Zea mays* L.). J. Exp. Bot. 62:2309–2318. doi:10.1093/jxb/erq373

Chander, S., Y. Meng, Y. Zhang, J. Yan, and J. Li. 2008. Comparison of nutritional traits variability in selected eighty-seven inbreds from Chinese maize (*Zea mays* L.) germplasm. J. Agric. Food Chem. 56:6506–6511. doi:10.1021/jf7037967

Commission International de l'Eclairage (CIE). 2007. Part 4: CIE 1976 L*a*b* colour space. CIE standard S014-4/E:2007. CIE, Vienna, Austria.

Darrigues, A., J. Hall, E. van der Knaap, and D.M. Francis. 2008. Tomato analyzer-color test: A new tool for efficient digital phenotyping. J. Am. Soc. Hortic. Sci. 133:579–586.

Flint-Garcia, S.A., A.L. Bodnar, and M.P. Scott. 2009. Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. Theor. Appl. Genet. 119:129–142. doi:10.1007/s00122-009-1115-1

Floyd, C.D., L.W. Rooney, and A.J. Bockholt. 1995. Measuring desirable and undesirable color in white and yellow food corn. Cereal Chem. 72:488–490.

Fox, G., and M. Manley. 2009. Hardness methods for testing maize kernels. J. Agric. Food Chem. 57:5647–5657. doi:10.1021/jf900623w

Fufa, H., G. Akalu, A. Wondimu, S. Taffesse, T. Gebre, K. Schlosser, H. Noetzold, and T. Henle. 2003. Assessment of protein nutritional quality and effects of traditional processes: A comparison between Ethiopian quality protein maize and five Ethiopian adapted normal maize cultivars. Nahrung/Food 47:269–273.

Gerdes, J.T., W.F. Tracy, and J.G. Coors. 1993. Compilation of North American maize breeding germplasm. Crop Science Society of America, Madison, WI.

Goggi, A.S., P. Caragea, L. Pollak, G. McAndrews, M. DeVries, and K. Montgomery. 2008. Seed quality assurance in maize breeding programs: Tests to explain variation in maize inbreds and populations. Agron. J. 100:337–343. doi:10.2134/agrojnl2007.0151

Hacisalihoglu, G., B. Larbi, and M. Settles. 2010. Near-infrared reflectance spectroscopy predicts protein, starch, and seed weight in intact seeds of common bean (*Phaseolus vulgaris* L.). J. Agric. Food Chem. 58:702–706. doi:10.1021/jf9019294

Harrigan, G.G., D. Lundry, S. Drury, K. Berman, S.G. Riordan, M.A. Nemeth, W.P. Ridley, and K.C. Glenn. 2010. Natural variation in crop composition and the impact of transgenesis. Nat. Biotechnol. 28:402–404. doi:10.1038/nbt0510-402

Igathinathane, C., L.O. Pordesimo, E.P. Columbus, W.D. Batcher, and S.R. Methuku. 2008. Shape identification and particle size distribution from basic shape parameters using ImageJ. Comput. Electron. Agric. 63:168–182. doi:10.1016/j.compag.2008.02.007

Itle, R.A., and E.A. Kabelka. 2009. Correlation between L*a*b* color space values and carotenoid content in pumpkins and squash (*Cucurbita* spp.). HortScience 44:633–637.

Jamal, A.Y.-S.M., and M.Z. Abdin. 2010. Sulfur-a general overview and interactions with nitrogen. Aust. J. Crop Sci. 4:523–529.

Jaradat, A.A., W. Goldstein, and K. Dashiell. 2010. Phenotypic structures and breeding value of open-pollinated corn varietal hybrids. Int. J. Plant Breed. 4:37–46.

Kaur, S., N. Singh, and J.C. Rana. 2010. *Amaranthus hypondriacus* and *Amaranthus caudatus* germplasm: Characteristics of plants, grain, and flours. Food Chem. 123:1227–1234. doi:10.1016/j.foodchem.2010.05.091

Kuhnen, S., P.M.M. Lemos, L.H. Capestrini, J.B. Ogliari, P.F. Dias, and M. Maraschin. 2011. Carotenoid and anthocyanin contents of grains of Brazilian maize landraces. J. Sci. Food Agric. 91:1548–1553. doi:10.1002/jsfa.4346

Miao, Y., D.J. Mulla, P.C. Robert, and J.A. Hernandez. 2006. Within-field variation in corn yield and grain quality responses to nitrogen fertilization and hybrid selection. Agron. J. 98:129–140. doi:10.2134/agronj2005.0120

Mittelmann, A., J.B. Filho, L.L. Nass, G.J.M. de Lima, C. Hara-Klein, and R.M. da Silva. 2011. Quantitative variation for grain quality in Brazilian maize populations. Sci. Agric. 68:50–56. doi:10.1590/S0103-90162011000100008

Moore, S.M., K.J. Stadler, D.C. Beitz, C.H. Stahl, W.A. Fithian, and K. Bregendahl. 2008. The correlation of chemical and physical corn kernel traits with production performance in broiler chickens and laying hens. Poult. Sci. 87:665–676. doi:10.3382/ps.2007-00184

Murphy, K.M., P.G. Reeves, and S.S. Jones. 2008. Relationship between yield and mineral nutrient concentrations in historical and modern spring wheat cultivars. Euphytica 163:381–390. doi:10.1007/s10681-008-9681-x

Nuss, E.T., and S.A. Tanumihardjo. 2010. Maize: A paramount staple crop in the context of global nutrition. Compr. Rev. Food Sci. Food Saf. 9:417–436.

Olsen, M.S., T.L. Krone, and R.L. Phillips. 2003. BSSS53 as a donor source for increased whole-kernel methionine in maize: Selection and evaluation of high-methionine inbreds and hybrids. Crop Sci. 43:1634–1642. doi:10.2135/cropsci2003.1634

Orak, H.H., I. Baser, K.Z. Korkut, O. Bilgin, and A. Orak. 2010. Investigation of peroxidase and lipoxygenase enzyme activities of dent corn (*Zea mays* L.) hybrids and their relationships with color and chemical characteristics. J. Agron. Crop Sci. 1:33–40.

Ortiz, R., S. Taba, V.H.C. Tovar, M. Mezzalama, Y. Xu, J. Yan, and J.H. Crouch. 2010. Conserving and enhancing maize genetic resources as global public goods- a perspective from CIMMYT. Crop Sci. 50:13–28. doi:10.2135/cropsci2009.06.0297

Ortiz-Monasterio, J.I., N. Palacios-Rojas, E. Meng, K. Pixley, R. Trethowan, and R.J. Pena. 2007. Enhancing the mineral and vitamin content of wheat and maize through plant breeding. J. Cereal Sci. 46:293–307. doi:10.1016/j.jcs.2007.06.005

Osborne, B.G. 2001. Application of near infrared spectroscopy in quality screening of early-generation material in cereal breeding programs. J. Near Infrared Spectrosc. 14:93–101. doi:10.1255/jnirs.595

Payne, R.W., S.A. Harding, D.A. Murray, D.M. Soutar, D.B. Baired, S.J. Welham, A.F. Kane, A.R. Gilmore, R. Thompson, R. Webster, and G.T. Wilson. 2007. The guide to GenStat release 10. Part 2: Statistics. VSN Int., Hemel Hempstead, UK.

Prandini, A., S. Sigolo, M. Morlacchini, A. Marocco, and M. Lo Ointo. 2011. High-protein maize in diets of growing broilers. Ital. J. Anim. Sci. doi:10.4081/ijas.2011.e55

Prasanna, B.M., S.K. Vasal, B. Kassahun, and N.N. Singh. 2001. Quality protein maize. Curr. Sci. 81:1308–1319.

Rasband, W.S. 2007. ImageJ: Image processing and analysis in Java. U.S. National Institute of Health, Bethesda, MD. http://rsb.info.nih.gov/ij/index.html (accessed 10 Apr. 2012).

Reeves, T.G., and K. Cassady. 2002. History and past achievements of plant breeding. Aust. J. Agric. Res. 53:851–863. doi:10.1071/AR02038

Rosales, A., L. Galicia, E. Oviedo, C. Islas, and N. Palacios-Rojas. 2011. Near-infrared reflectance spectroscopy (NIRS) for protein, tryptophan, and lysine evaluation in quality protein maize (QPM) breeding programs. J. Agric. Food Chem. 59:10781–10786. doi:10.1021/jf201468x

Salhuana, W., L.M. Pollak, M. Ferrer, O. Paratori, and G. Vivo. 1998. Agronomic evaluation of maize accessions from Argentina, Chile, USA, and Uruguay. Crop Sci. 38:866–872. doi:10.2135/cropsci1998.0011183X003800030040x

Scott, M.P., and M. Blanco. 2009. Evaluation of the grain methionine, lysine, and tryptophan contents of maize (*Zea mays* L.) germplasm in the germplasm enhancement of maize project. Plant Genet. Resour. 7:237–243. doi:10.1017/S1479262109264148

Scott, M.P., A. Darrigues, T.S. Stahly, and K. Lamkey. 2008. Recurrent selection to alter grain methionine concentration and improve nutritional value of maize. Crop Sci. 48:1705–1713. doi:10.2135/cropsci2008.01.0010

Seebauer, J.R., G.W. Singletary, P.M. Krumpelman, M.L. Ruffo, and F.E. Below. 2010. Relationship of source and sink in determining kernel composition of maize. J. Exp. Bot. 61:511–519. doi:10.1093/jxb/erp324

Shiferaw, B., B.M. Prasanna, J. Hellin, and M. Bänziger. 2011. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. Food Sci. 3:307–327. doi:10.1007/s12571-011-0140-5

Singh, A.K., T.K. Coleman, M.T. Tollenaar, and E.A. Lee. 2011. Nature of the genetic variation in an elite maize breeding cross. Crop Sci. 51:75–83. doi:10.2135/cropsci2010.06.0378

Skogerson, K., G.G. Harrigan, T.L. Reynolds, S.C. Halls, M. Ruebelt, A. Ionadolino, A. Pandravada, K.C. Glenn, and O. Fiehn. 2010. Impact of genetics and environment on the metabolite composition of maize grain. J. Agric. Food Chem. 58:3600–3610. doi:10.1021/jf903705y

Spielbauer, G., P. Armstrong, J.W. Baier, W.B. Allen, K. Richardson, B. Shen, and A.M. Settles. 2009. High-throughput near-infrared reflectance spectroscopy for predicting quantitative and qualitative composition phenotypes of individual maize kernels. Cereal Chem. 86:556–564. doi:10.1094/CCHEM-86-5-0556

StatSoft. 2011. STATISTICA (data analysis software system). Version 10. StatSoft, Inc., Tulsa, OK. www.statsoft.com (accessed 1 Apr. 2012).

Taboada-Gaytan, O., L.M. Pollak, L.A. Johnson, S.R. Fox, and K.T. Montgomery. 2010. Variation among physical, compositional, and wet-milling characteristics of the F_1 generation of corn hybrids of introgressed exotic and adapted inbred lines. Cereal Chem. 87:175–181. doi:10.1094/CCHEM-87-3-0175

Thomison, P.R., D.J. Barker, A.B. Geyer, L.D. Lotz, H.J. Siegrist, and T.L. Dobbles. 2003. Amino acid composition of topcross high-oil maize grain. Plant Genet. Resour. 1:89–95.

USDA-Agricultural Marketing Service (USDA-AMS). 2013a. Copy of plant variety protection certificate for LH119. USDA-AMS, Beltsville, MD. http://apps.ams.usda.gov/CMS//AdobeImages/008200064.pdf (accessed 1 Mar. 2011).

USDA-Agricultural Marketing Service (USDA-AMS). 2013b. Copy of plant variety protection certificate for LH82. USDA-AMS, Beltsville, MD. http://apps.ams.usda.gov/CMS//AdobeImages/008500037.pdf (accessed 1 Mar. 2011).

USDA-ARS. 2013a. Ames 20120. USDA-ARS, Germplasm Resource Information Network, Beltsville, MD. www.ars-grin.gov/cgi-bin/npgs/acc/display.pl?1084203 (accessed 1 Feb. 2008).

USDA-ARS. 2013b. GEM germplasm releases and key traits. North Central Regional Plant Introduction Station, Ames, IA. www.public.iastate.edu/~usda-gem/Official_Documents/GEM_Germplasm_Releases.htm (accessed 1 Dec. 2009).

USDA-ARS. 2013c. Image for: PI 467165 – Sample of bulk kernels. website. USDA-ARS, Germplasm Resource Information Network, Beltsville, MD. www.ars-grin.gov/cgi-bin/npgs/html/dispimage.pl?38396 (accessed 1 Feb. 2008).

USEPA. 2007. SW-846 test methods. USEPA, Washington, DC. www.epa.gov/osw/hazard/testmethods/sw846/index.htm (accessed 4 Feb. 2012).

Wang, L., Q. Wang, H. Liu, L. Liu, and Y. Du. 2012. Determining the contents of protein and amino acids in peanuts using near-infrared reflectance spectroscopy. J. Sci. Food Agric. doi:10.1002/jsfa.5738

Welch, R.M., and D. Graham. 2004. Breeding for micronutrients in stable food crops from a human nutrition perspective. J. Exp. Bot. 55:353–364. doi:10.1093/jxb/erh064

Yeh, F.C., R.-C. Yang, T.B.J. Boyle, Z.-H. Ye, and J.X. Mao. 2000. POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and biotechnology Center, University of Alberta, Edmonton, AB, Canada.

Zar, J.H. 1996. Biostatistical analysis, 3rd ed. Prentice Hall, Englewood Cliffs, NJ.