Institute of

Plant Breeding, Seed Science and Population Genetics

University of Hohenheim

Institute of Plant Breeding

Prof. Dr. A. E. Melchinger

STRATEGIES FOR SELECTING HIGH-YIELDING AND BROADLY ADAPTED MAIZE HYBRIDS FOR THE TARGET ENVIRONMENT IN EASTERN AND SOUTHERN AFRICA

Dissertation

Submitted in fulfillment of the requirements for the degree

“Doktor der Agrarwissenschaften“

(Dr. sc. Agr. / *Ph. D.* in Agricultural Sciences)

to the

Faculty of Agricultural Sciences

presented by

Master of Science ETH in Agroecosystem Science

Sandra VANESSA WINDHAUSEN née Weber

from Marburg

2012

This thesis was accepted as a doctoral dissertation in fulfillment of the requirements for the degree “Doktor der Agrarwissenschaften” (Dr. sc. Agr. / Ph. D. in Agricultural Sciences) by the Faculty of Agricultural Sciences at the University of Hohenheim, on October 22, 2012.

Day of oral examination: December 4 2012

Examination Committee

Vice-Dean and Head of the Committee: Prof. Dr. S. Böttinger

Supervisor and Reviewer: Prof. Dr. A. E. Melchinger

Co-reviewer: Prof. Dr. J.-L. Araus Ortega

Additional examiner: Prof. Dr. H.-P. Piepho

Table of contents

---

[1] Weber VS, Araus JL, Cairns JE, Sanchez C, Melchinger AE, Orsini E (2012): Prediction of grain yield using reflectance spectra of canopy and leaves in maize plants grown under different water regimes. Field Crops Research 128: 82-90.

[2] Weber VS, Melchinger AE, Magorokosho C, Makumbi D, Bänziger M, Atlin GN (2012): Efficiency of Managed-Stress Screening of Elite Maize Hybrids under Drought and Low Nitrogen for yield under Rainfed Conditions in Southern Africa. Crop Science 52: 1011-1020

[3] Windhausen VS, Wagener S, Magorokosho C, Makumbi D, Vivek B, Piepho H-P, Melchinger AE, Atlin GN (2012): Strategies to Subdivide a Target Population of Environments: Results from the CIMMYT-led Maize Hybrid Testing Programs in Africa. Crop Science 52: 2143-2152

[4] Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink J-L, Sorrels M, Raman B, Cairns JE, Tarekegne A, Semagn K, Beyene Y, Grudloyma P, Technow F, Riedelsheimer C, Melchinger AE (2012): Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and non-overlapping environments. G3, 101534/g3112003699: 1427-1436

1. General introduction

Maize (*Zea mays L.*) is grown on 25 million hectares in sub-Saharan Africa by smallholder farmers primarily for human consumption and livestock feed (Shiferaw et al., 2011). Yields in Africa are considerably lower than world average, because maize crops are mostly grown in drought and low-N environments (FAO, 2010). Projections of decreasing precipitation (Williams and Funk, 2010), increasing fertilizer prices and expansion of maize cultivation into less fertile lands (IPCC, 2007) further accentuate the need to provide farmers with drought and low-N tolerant maize varieties.

Efficiency of direct, indirect and index selection

Genetic improvement in maize can be achieved through (i) direct selection of grain yield in the target environment, (ii) indirectly for a secondary trait or grain yield in another test environment, or (iii) through index selection using information from both the direct and indirect test environment (Byrne et al., 1995). The gain from direct selection can be estimated as:

$$R_x = i_x \sqrt{h_x^2}\sigma_g \, , \qquad [1]$$

where i is selection intensity defined as the fraction of the current population retained to be used as parents for the next generation, h^2 is the heritability of a trait that is defined as the ratio between the genetic and phenotypic variance, and σ_g the square root of the genetic variance (Falconer and Mackay, 1996). Selection gain per year (y) can be estimated by extending equation [1] by the factor of the parental control (c):

$$R_x / y = c i_x \sqrt{h_x^2}\sigma_g \qquad [2]$$

The degree of parental control depends on whether selection is conducted before or after anthesis. At harvest, selection is only being made among females implying that males with low performance contributed to the pollen cloud and thus to the next generation (c=0.5). Selection gain can be doubled by selecting female and male parents for crossing prior to anthesis (c=1).

Gain from indirect selection can be predicted based on the genetic correlation (r_g) between trait of interest (x) and the trait used for selection (y):

$$CR_x = i_y \sqrt{h_y^2} r_g \sigma_g \quad [3]$$

Assuming the same selection intensity for direct and indirect selection, the efficiency of indirect selection will be higher than that of indirect selection if the heritability of indirect selection is higher and if the genetic correlation between indirect and direct selection is high:

$$\frac{CR_x}{R_x} = \frac{i_y r_g}{i_x} \sqrt{\frac{h_y^2}{h_x^2}} \quad [4]$$

With direct and indirect selection, the economic weights of the genotypic performance of a genotype for each test environment or trait are not considered. In contrast, index selection demands appropriate weighting of both strategies (Mistele et al., 1994). Smith (1936) and Hazel (1943) showed that the unknown index weights (*b*) can be derived by multiplying the inverse of the phenotypic variance-covariance matrix (*P*), the genetic variance-covariance matrix (*G*) and the economic weights (*a*) of each trait:

$$b = P^{-1} Ga \quad [5]$$

The economic weight of a targeted test environment or trait can be considered as 1 and that of another test environment or trait as 0 to optimize recommendation for the target environment. Gain of index selection using information from both test environments or traits can be expressed as (Wricke and Weber, 1986):

$$R = i \frac{b'G}{\sqrt{b'Pb}} \quad [6]$$

Indirect selection for maize grain yield using a novel secondary trait

The ability to accurately predict grain yield at anthesis by using secondary traits would be useful to select superior maize lines for crossing, thereby increasing the selection intensity and parental control whilst reducing phenotyping costs within the early stages of a breeding program. For a secondary trait to be useful, it should be (i) genetically variable and correlated with grain yield, (ii) less affected by the environment and higher heritable than grain yield, (iii) cheaper and/or faster to measure than grain yield, and (iv) able to provide an estimate of grain yield potential before harvest (Edmeades et al., 1996; Araus et al., 2008). Anthesis-silking interval is perhaps the most prominent example of an effective secondary trait (Bolaños and Edmeades, 1993; Edmeades et al., 1999). In elite maize germplasm breeders have reduced the anthesis-silking interval and the variation in the trait considerably so that further gains from reduction of its value may be less likely (Byrne et al., 1995; Monneveux et al., 2008). Thus, novel secondary traits need to be identified to support selection of high-yielding and stress tolerant genotypes. The measurement of leaf and canopy reflectance may be used to predict grain yield and traits related to high photosynthetic capacity, above ground biomass and plant water content in future environments (Aparicio et al., 2000; Royo et al., 2003). Prediction of grain yield was reported to be stronger and more robust when the whole spectra was used for prediction than with previously assayed spectral indices developed on the basis of ratios or differences in the reflectance at a given wavelength (Ferrio et al., 2004, 2005). Utilization of the complete spectra for the prediction of genotype performance requires the development of calibration models that relate the spectra to the trait of interest. In the process of calibration development, it has to be defined at which plant developmental stage (*i.e.*, at anthesis, grain filling), measurement level (*i.e.*, at leaf or canopy level), and in which test environment (*i.e.*, managed stress or optimal) the highest percentage of genetic variance is explained.

Definition of the target and test environment

For a variety to become popular among farmers, it must combine tolerance to random abiotic stress environments, including drought and low-N stress, with high grain yield potential in optimal environments (Bänziger et al., 1999; Kumar et al., 2008). Selection under random abiotic stress is made difficult by its unpredictable occurrence and intensity. To select genotypes tolerant to random abiotic stress, the maize breeding program of the International Maize and Wheat Improvement Center (CIMMYT) selects for grain yield and several secondary traits in managed drought and low-N environments in all replicated yield testing stages (Bänziger et al., 2000). Selection under target stresses accelerate breeding gains, if these represent the growing conditions in farmers‘ fields implying a moderate to high genetic correlation between test and target environment (Atlin and Frey, 1990; Ceccarelli et al., 1992; Bänziger et al., 1997, 2006). Managed drought trials are conducted in the dry season, where day length, humidity, and disease pressure may differ from the main growing season. Furthermore, drought stress is imposed during anthesis and grain-filling by withholding irrigation and low N stress is imposed by planting trials without fertilization in N depleted fields causing more than 70% grain yield reduction (Bänziger et al., 1997). Maize yields in farmers’ fields are rarely reduced by a defined stress at a certain plant growth stage but rather by a combination of stresses occurring at the same or different growth stages. As a consequence, selection in managed drought or low-N environments might not necessarily result in varieties adapted to random abiotic stress. Alternatively, genotypes can be evaluated in optimal, high-yielding environments wherein the testing effort and heritability of grain yield are usually higher than in low-yielding, managed or random abiotic stress environments. As such, indirect selection in high-yielding environments may be more efficient than direct selection in low-yielding environments if the genetic correlation among test environments is high. A combined selection across optimal, managed drought, low-N and random abiotic stress conditions would be ideal as it would be conducted across the growing conditions occurring in farmers’ fields. Nevertheless, combined selection across the undivided target environment would only be appropriate if the evaluation across different yield levels does not involve genotype-by-environment interactions (Atlin et al., 2000a; b).

Evaluation of pros and cons for subdividing the target environment

The existence of genotype-by-environment interactions may imply that the best genotype in one environment may not be the best in another. To cope with large genotype-by-environment interactions associated with consistent subgroupings of environments within the target environment, the target environment can be subdivided into several smaller and more homogeneous subregions. For example, CIMMYT currently subdivides the target environment in Eastern and Southern Africa according to geographic (*i.e.*, Eastern and Southern African subregions), climate (*i.e.*, tropical and subtropical subregions) or elevation differences (*i.e.*, lowland and mid-altitude subregions, Bänziger et al., 2004, 2006). When a target environment is subdivided, it is implicitly assumed that consistent genotype-by-subregion interactions exist. Examination of variance components can provide initial information about the magnitude and practical importance of genotype-by-subregion interaction variance. If the genotype-by-subregion interaction variance is small relative to the genetic variance, this indicates that there is little specific adaptation to the respective subregions and that simultaneous selection for a wide range of environments is possible and may be cost effective because the cost of breeding is spread across a larger market (Atlin et al., 2000b). If genotype-by-subregion interactions are repeatable and the genetic correlations among subregions are low, this indicates that greater gains from selection may result by subdividing the target environment. In this case, the increase in genetic variance by converting the genotype-by-subregion variance into the genetic variance needs to counterbalance the lower heritability of genotype means associated with the reduced testing effort within the smaller subregions (Atlin et al., 2000a, 2001; Piepho and Möhring, 2005). Even if there is a consistent subregion effect, information of all subregions can be used by applying index selection and computing a weighted mean of mean yields in the targeted and neighboring subregions, with weights depending on the similarity between subregions and the number of trials per subregion (Piepho and Möhring, 2005). The resulting estimator is beneficial when selecting for global adaptation, minimizing prediction errors and maximizing the expected gain from selection. Selection gain per year may then be further increased by increasing the selection intensity or by acceleration of the breeding cycle (Falconer and Mackay, 1996).

Support of genotypic selection using genomic prediction

In maize, selection for testcross performance requires at least two years. The success of selecting genotypes to be parents of the next generation according to their phenotypic value may be limited if the degree of correspondence between phenotypic and breeding values is low. As the breeding value of a genotype is equal to the sum of average effects of the genes it carries, it would be advantageous to support genotype selection and accelerate the breeding cycle by the use of molecular markers. Genomic prediction has been proposed to predict genotype performance by estimating all marker effects simultaneously based on a training population of individuals with both phenotypic and genotypic data (Meuwissen et al., 2001; Heffner et al., 2009). Subsequently, estimated marker effects are used to predict performance of a validation set having only genotypic data. On the basis of predicted performance, genotypes can then be selected for advancement in the breeding cycle without phenotypic evaluation which would increase the selection intensity tremendously. The cumulative gain from three cycles of genomic prediction approaches 1.5 times the gain from one cycle of phenotypic selection if prediction accuracies are equal or greater than half of the square root of the heritability of the phenotyped trait (Lorenzana and Bernardo, 2009). Even higher levels of prediction accuracy were reported for maize grain and biomass yield (Crossa et al., 2010; Albrecht et al., 2011; Riedelsheimer et al., 2012; Zhao et al., 2012). Before incorporating genomic prediction in a plant breeding program, one has to clearly define the breeding scenario in which genomic prediction will be applied, *i.e.*, whether one wants to predict performance within or across breeding populations. For each scenario it has to be assessed if high prediction accuracies reported for maize yield in recent studies hold up regarding the heritability of the trait of interest, the sample size of the training set, the number of test environments, the population structure within the training and validation sets and the genetic relationship between training and validation sets.

Given the proposed strategies for improving performance in the target environment, the breeder needs to evaluate how they could be combined to achieve highest gains in the breeding program. Consequently, the objectives of my thesis were to

(1) evaluate the potential of leaf and canopy spectral reflectance as novel secondary traits to predict grain yield across different environments,
(2) estimate to what extent indirect selection in managed drought and low-N environments is predictive of grain yield in random abiotic stress environments,
(3) investigate whether subdividing the target environment into climate, altitude geographic, yield level or country subregions is likely to increase rates of genetic gain, and
(4) evaluate the prospects to perform genomic prediction in the presence of population structure

References

Albrecht, T., V. Wimmer, H.-J. Auinger, M. Erbe, C. Knaak, M. Ouzunova, H. Simianer, and C.-C. Schön. 2011. Genome-based prediction of testcross values in maize. TAG 123(2): 339-50.

Aparicio, N., D. Villegas, J. Casadesus, J.L. Araus, and C. Royo. 2000. Spectral vegetation indices as nondestructive tools for determining durum wheat yield. Agronomy Journal 92(1): 83-91.

Araus, J.L., G.A. Slafer, C. Royo, and M.D. Serret. 2008. Breeding for yield potential and stress adaptation in cereals. Critical Reviews in Plant Sciences 27(6): 377-412.

Atlin, G.N., R.J. Baker, K.B. McRae, and X. Lu. 2000a. Selection Response in Subdivided Target Regions. Crop Sci. 40: 7-13.

Atlin, G.N., M. Cooper, and Å. Bjørnstad. 2001. A comparison of formal and participatory breeding approaches using selection theory. Euphytica 122(3): 463-475.

Atlin, G.N., and K.J. Frey. 1990. Selecting oat lines for yield in low-productivity environments. Crop Science 30(3): 556-561.

Atlin, G.N., K.B. McRae, and X. Lu. 2000b. Genotype x region interaction for two-row barley yield in Canada. Crop Science 40(1): 1-6.

Bolaños, J., and G.O. Edmeades. 1993. Eight cycles of selection for drought tolerance in lowland tropical maize. I. Responses in grain yield, biomass, and radiation utilization. Field Crops Research 31(3-4): 253-268.

Byrne, P.F., J. Bolaños, G.O. Edmeades, and D.L. Eaton. 1995. Gains from Selection under Drought versus Multilocation Tropical Maize Populations Testing in Related. Crop Science 35: 63-69.

Bänziger, M., F.J. Betran, and H.R. Lafitte. 1997. Efficiency of high-nitrogen selection environments for improving maize for low-nitrogen target environments. Crop Science 37(4): 1103-1109.

Bänziger, M., G.O. Edmeades, D. Beck, and M. Bellon. 2000. Breeding for Drought and Nitrogen Stress Tolerance in Maize: From Theory to Practice. CIMMYT, Mexico, D.F.

Bänziger, M., G.O. Edmeades, and H.R. Lafitte. 1999. Selection for drought tolerance increases maize yields across a range of nitrogen levels. Crop Science 39(4): 1035-1040.

Bänziger, M., P.S. Setimela, D. Hodson, and B. Vivek. 2004. Breeding for improved drought tolerance in maize adapted to Sub-Saharan Africa. p. 1-10. *In* Proceedings of the 4th international Crop Science Congress. 26. Sep-1.Oct 2004, Brisbane, Australia.

Bänziger, M., P.S. Setimela, D. Hodson, B. Vivek, and M. Banziger. 2006. Breeding for improved abiotic stress tolerance in maize adapted to southern Africa. Agricultural Water Management 80(1-3): 212-224.

Ceccarelli, S., S. Grando, and J. Hamblin. 1992. Relationship between barley grain yield measured in low- and high-yielding environments. Euphytica 64(1-2): 49-58.

Crossa, J., G. De Los Campos, P. Pérez, D. Gianola, J. Burgueño, J.L. Araus, D. Makumbi, R.P. Singh, S. Dreisigacker, J. Yan, V. Arief, M. Banziger, and H.-J. Braun. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186(2): 713-724.

Edmeades, G.O., J. Bolaños, and S.C. Chapman. 1996. Value of Secondary Traits in Selecting for Drought Tolerance in Tropical Maize. *In* Edmeades, G.O., Bänziger, M., Mickelson, H., Peña-Valdivia, C.B. (eds.), Developing Drought- and Low N-Tolerant Maize. CIMMYT, Mexico, D.F.

Edmeades, G.O., J. Bolaños, S.C. Chapman, H.R. Lafitte, and M. Bänziger. 1999. Selection improves drought tolerance in tropical maize populations: I. Gains in biomass, grain yield, harvest index. Crop Science 39(5): 1306-1315.

FAO. 2010. The State of Food Insecurity in the World. Addressing food insecurity in protracted crises. (FAO, Ed.). 19.03.2011, Food and Agriculture Organization of the United Nations, http://www.fao.org/publications/sofi/en/.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics. Pearson Education Limited, Essex, England.

Ferrio, J.P., E. Bertran, M.M.   , and J.L. Araus. 2004. Estimation of grain yield by near-infrared reflectance spectroscopy in durum wheat. Euphytica 137(3): 373-380.

Ferrio, J.P., D. Villegas, J. Zarco, N. Aparicio, J.L. Araus, and C. Royo. 2005. Assessment of durum wheat yield using visible and near-infrared reflectance spectra of canopies. Field Crops Research 94(2-3): 126-148.

Heffner, E.L., M.E. Sorrells, and J.L. Jannink. 2009. Genomic Selection for Crop Improvement. Crop Sci. 49: 1-12.

IPCC. 2007. Contribution of Working Group II to the Forth Assessment Report of the Intergovernmental Panel on Climate Change. p. 976. *In* IPCC (ed.), Climate Change 2007: Impacts, Adaptation and Vulnerability. Cambridge University Press, Cambridge, UK.

Kumar, A., J. Bernier, S. Verulkar, H.R. Lafitte, and G.N. Atlin. 2008. Breeding for drought tolerance: Direct selection for yield, response to selection and use of drought-tolerant donors in upland and lowland-adapted populations. Field Crops Research 107(3): 221-231.

Lorenzana, R.E., and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. TAG 120(1): 151-61.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157: 1819-1829.

Monneveux, P., C. Sanchez, and A. Tiessen. 2008. Future progress in drought tolerance in maize needs new secondary traits and cross combinations. Journal of Agricultural Science 146(3): 287-300.

Piepho, H.P., and J. Möhring. 2005. Best linear unbiased prediction of cultivar effects for subdivided target regions. Crop Science 45(3): 1151-1159.

Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, R. Sulpice, T. Altmann, M. Stitt, L. Willmitzer, and A.E. Melchinger. 2012. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nature Genetics 44(2): 217-20.

Royo, C., N. Aparicio, D. Villegas, J. Casadesus, P. Monneveux, and J.L. Araus. 2003. Usefulness of spectral reflectance indices as durum wheat yield predictors under contrasting Mediterranean conditions. International Journal of Remote Sensing 24(22): 4403-4419.

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 Security 3(3): 307-327.

Williams, A.P., and C. Funk. 2010. A westward extension of the warm pool leads to a westward extension of the Walker circulation, drying eastern Africa. Climate Dynamics: 1-19.

Wricke, G., and W.E. Weber. 1986. Quantitative Genetics and Selection in Plant Breeding. Walter de Gruyter, Berlin, New York.

Zhao, Y., M. Gowda, W. Liu, T. Würschum, H.P. Maurer, F.H. Longin, N. Ranc, and J.C. Reif. 2012. Accuracy of genomic selection in European maize elite breeding populations. Theoretical and Applied Genetics 124(4): 769-76.

2. Prediction of grain yield using reflectance spectra of canopy and leaves in maize plants grown under different water regimes

Weber VS[1], Araus J-L[2,3], Cairns JE[3], Sanchez C[3], Melchinger AE[1], and Orsini E[1]

[1]Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Germany
[2]Unitat de Fisiologia Vegetal, Facultat de Biologia, Universitat de Barcelona, Spain
[3]International Maize and Wheat Improvement Center (CIMMYT), Mexico

Field Crops Research 128 (2012) 82–90

doi:10.1016/j.fcr.2011.12.016

The original publication is available at http://www.elsevier.com/locate/fcr

Abstract

The ability to accurately estimate grain yield using spectral reflectance measurements prior harvest could be used to reduce phenotyping time and costs. In this study, grain yield of 300 maize testcrosses grown under different water and temperature regimes in the dry season 2010 was predicted using spectral reflectance (495–1853 nm) of both leaves and canopy measured between tassel emergence until milk-grain stage. Partial least square regression (PLSR) was used for data analysis. Coefficients of determination (R^2) between predicted and actual grain yield were highest for measurements conducted at anthesis and milk-grain stage, explaining at maximum 23% and 40% of the genotypic variation in grain yield after validation, respectively. PLSR models explained a higher proportion of the genetic variation in grain yield under drought stress compared to well-watered conditions. The association between predicted and actual grain yield was stronger in spectral reflectance measurements taken at the leaf level compared to canopy level. By combining the most predictive PLSR models across trials, at maximum of 40% of the variation in grain yield could be explained in each trial with a relative efficiency of selection of 0.88 and 0.68 using leaf and canopy reflectance, respectively. The most relevant wavelengths for predicting grain yield were associated with photosynthetic capacity (495–680 nm), red inflection point (680–780 nm) and plant water status (900, 970, and 1450 nm, 1150–1260 nm, and 1520–1540 nm). Additional wavelengths based on leaf (800, 1000, and 1260–1830 nm) and canopy (988–999 nm and 1430–1640 nm) reflectance of unknown physiological relevance were also identified for prediction of grain yield. Caution must be exercised before integrating our spectral reflectance approach into a breeding program because this is a pilot study based on a single location and season.

3. Efficiency of Managed-Stress Screening of Elite Maize Hybrids under Drought and Low Nitrogen for Yield under Rainfed Conditions in Southern Africa

Weber VS[1], Melchinger AE[1], Magorokosho C[2], Makumbi D[2], Bänziger M[2], and Atlin GN[2]

[1]Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Germany
[2]International Maize and Wheat Improvement Center (CIMMYT), Mexico.

Crop Science 52 (2012) 1011–1020

DOI: 10.2135/cropsci2011.09.0486

Abstract

Maize yields in eastern and southern Africa are considerably lower than the world average because its cultivation is often prone to drought and low soil fertility. To select genotypes adapted to these conditions, CIMMYT selects indirectly under managed drought and low-N stress, although the selection efficiency of this approach is not known. A retrospective analysis of 704 elite hybrid trials conducted from 2001 to 2009 was used to evaluate the relative ability of optimal, low-N, and managed drought trials to predict performance under random abiotic stress and low-N conditions usually faced by African farmers. Well-fertilized trials conducted in the rainy season were categorized as having experienced random abiotic stress if mean yield was below 3 t ha^{-1} and the yield-anthesis date correlation was below 0.1; otherwise they were categorized as optimal. Genetic correlations were highest between random abiotic stress and low-N or optimal conditions. Heritability was highest under optimal and lowest in random abiotic stress conditions. Indirect selection under low-N and optimal conditions was predicted to be more efficient than direct selection under random abiotic stress. In contrast, indirect selection under managed drought conditions was not efficient to predict genotype performance under random abiotic stress conditions, especially for early maturing genotypes. For predicting performance under low-N conditions direct selection was most efficient. Consequently, elite maize hybrids tolerant to random abiotic stress can be most efficiently selected under optimal and/or low-N conditions while low-N tolerant genotypes should be selected directly under low N.

4. Strategies to Subdivide a Target Population of Environments: Results from the CIMMYT-led Maize Hybrid Testing Programs in Africa

Windhausen VS[1], Wagener S[1], Magorokosho C[3], Makumbi D[3], Vivek B[3], Piepho H-P[2], Melchinger AE[1], and Atlin GN[3]

[1]Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Germany
[2]Institute of Crop Science, University of Hohenheim, Germany
[3]International Maize and Wheat Improvement Center (CIMMYT), Mexico.

Crop Science 52 (2012): 2143-2152

DOI: 10.2135/cropsci2012.02.0125

Windhausen et al. (2012), Crop Science 52: 2143-2152

Abstract

To develop stable and high-yielding maize hybrids for a diverse target environment, breeders have to evaluate whether greater gains result from selection across the undivided target environment or within more homogeneous and smaller subregions. Currently, CIMMYT subdivides the target environment in eastern and southern Africa into climatic and geographic subregions. To determine whether selection within subregions results in greater gains than selection across the undivided target environment, yield data of 448 maize hybrids evaluated in 513 trials across 17 countries from 2001 to 2009 were used. The trials were grouped according to five subdivision systems into climate, altitude, geographic, country, and yield-level subregions. For the first four subdivision systems, genotype-by-subregion interaction variance was low, suggesting broad adaptation of maize hybrids across eastern and southern Africa. In contrast, genotype-by-yield level interactions and moderate genetic correlations between low- and high-yielding subregions were observed. Therefore, hybrid means should be estimated separately for low- and high-yielding subregions considering the yield-level effect as fixed and appropriately weighting information from both subregions. This strategy was at least 10% better in terms of predicted gains than direct or indirect selection using only data from low- or high-yielding trials and should facilitate the identification of hybrids that perform well in both subregions.

5. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and non-overlapping environments

Windhausen VS[1], Atlin GN[2], Hickey JM[2], Crossa J[2], Jannink J-L[3], Sorrells M[3], Raman B[2], Cairns JE[2], Tarekegne A[2], Semagn K[2], Beyene Y[2], Grudloyma P[4], Technow F[1], Riedelsheimer C[1], and Melchinger AE[1]

[1]Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Germany
[2]International Maize and Wheat Improvement Center (CIMMYT), Mexico
[3]Department of Plant Breeding and Genetics, Cornell University, Ithaca 14853, NY, USA
[4]Nakhon Sawan Field Crops Research Center, Thailand

G3, 10.1534/g3.112.003699: 1427-1436

The original publication is freely available online at http://www.g3journal.org/

Abstract

Genomic prediction is expected to considerably increase genetic gains by increasing selection intensity and accelerating the breeding cycle. In this study, marker effects estimated in 255 diverse maize (*Zea mays L.*) hybrids were used to predict grain yield, anthesis date, and anthesis-silking interval within the diversity panel and testcross progenies of 30 F_2-derived lines from each of five populations. Although up to 25% of the genetic variance could be explained by cross validation within the diversity panel, the prediction of testcross performance of F_2-derived lines using marker effects estimated in the diversity panel was on average zero. Hybrids in the diversity panel could be grouped into eight breeding populations differing in mean performance. When performance was predicted separately for each breeding population on the basis of marker effects estimated in the other populations, predictive ability was low (*i.e.*, 0.12 for grain yield). These results suggest that prediction resulted mostly from differences in mean performance of the breeding populations and less from the relationship between the training and validation sets or linkage disequilibrium with causal variants underlying the predicted traits. Potential uses for genomic prediction in maize hybrid breeding are discussed emphasizing the need of (1) a clear definition of the breeding scenario in which genomic prediction should be applied (*i.e.*, prediction among or within populations), (2) a detailed analysis of the population structure before performing cross validation, and (3) larger training sets with strong genetic relationship to the validation set.

6. General discussion

The present study was based on three large data sets on maize hybrid performance. In the following I discuss, how selection gain of the CIMMYT maize breeding program could be increased regarding the definition of the test environments, the use of index selection, the implementation of genomic prediction, the support of selection using secondary traits, the allocation of testing resources, and the improvement of field designs.

Definition of the test environments

The maize breeding program of CIMMYT in Africa is currently subdivided into Eastern and Southern Africa. This subdivision of the target environment did not increase selection gain, because genotype-by-geographic region interaction variance was small relative to the genetic variance (Windhausen et al., 2012b). Consequently, simultaneous selection across Eastern and Southern Africa is possible and cost effective. This finding is in accordance with the fact that modern maize hybrids often exhibit very wide geographical adaptation (Braun et al., 2010). This implies that the Eastern and Southern African breeding programs of CIMMYT should be consolidated and that breeders need to increase their collaboration and the exchange of genotypes. Within the consolidated Eastern and Southern African breeding programs, selection decisions should be made separately for performance in low- and high-yielding environments, as the genotype-by-yield level interaction variance was as high as the genetic variance and the genetic correlations between low- and high-yielding environments were only moderate (Windhausen et al., 2012b).

As yields in farmers’ fields are usually reduced by a combination of different stresses occurring at the same or different plant growth stages, selection in low-yielding environments may be optimal to serve farmers needs. It is difficult to predict whether a field will be low-yielding before sowing which complicates the allocation of testing resources. Consequently, the question arises, whether it is possible to select indirectly for genotypes adapted to random abiotic stress occurring in low-yielding environments using managed stress and/or optimal test environments. Even though selection under target stresses may accelerate breeding gains, the difficulty of choosing appropriate test environments, given a highly variable target environment, may limit the identification of superior genotypes (Ceccarelli et al., 1992; Bänziger et al., 2000, 2006). A retrospective analysis across 9 years, over 600 trials and 448 advanced maize hybrids showed that the performance under random abiotic stress was most strongly correlated with that under low-N and optimal environments and less to that in managed drought stress environments (Weber et al., 2012b). Because indirect selection in optimal- or low-N environments implies a possibility of discarding some genotypes that may be high yielding under random abiotic stress, a combined evaluation across high- (*i.e.*, optimal) and low-yielding environments (*i.e.*, low-N, managed drought and random abiotic stress) might be of advantage.

Use of index selection

In the analysis of multi-location trials, random abiotic stress and optimal environments are usually not separated and considered as one target environment (Weber et al., 2012b). The disadvantage of estimating hybrid means ignoring genotype-by-yield level interactions is that effectively much greater weight is given to high- than to low-yielding trials, due to the higher testing effort and genetic variance (Windhausen et al., 2012b). This may explain why crop varieties bred primarily under high-yielding environments failed to have an impact in low-yielding production environments (Atlin and Frey, 1990; Ceccarelli et al., 1992; Atlin et al., 2001). The key to identify broadly-adapted genotypes for the target environment is the combined selection across low- and high-yielding environments considering the yield-level effect as fixed, thereby using information from both environments, appropriately weighted, in estimating hybrid performance within each environment (Windhausen et al., 2012b). The weights given to low- and high-yielding environments depend on the genetic variances within, the genetic correlation between, and the number of trials conducted within both environments (Piepho and Möhring, 2005). Combining information from low- and high-yielding environments has been shown to be beneficial for selecting cereal varieties adapted to organic farming conditions (Przystalski et al., 2008) and maize hybrids adapted to low-yielding environments (Windhausen et al., 2012b). Thus, index selection should be implemented into the CIMMYT maize breeding program for making breeding decisions, as it facilitates the identification of hybrids that perform well in low- and high-yielding environments and further has the potential to be extended to multiple traits (*i.e.*, actual or predicted grain yield, anthesis date or other secondary traits).

Implementation of genomic prediction

Genomic prediction is expected to considerably increase genetic gains by increasing selection intensity and accelerating the breeding cycle. Based on two large data sets on the performance of maize hybrids it was shown that prediction resulted mostly from differences in mean performance of the breeding populations (Windhausen et al., 2012a). Owing to differences in linkage disequilibrium and linkage phases between different breeding populations of CIMMYT maize germplasm, marker effects estimated in one breeding population could not be easily transferred to another, which resulted in low predictive abilities. The same was true when marker effects estimated in a diversity panel of 255 maize hybrids were used to predict performance of testcross progenies of 30 F_2-derived lines from each of five populations generated using nine lines, four of which were part of the diversity panel. Similar results were reported for 16 economically important traits in US Angus beef cattle (Saatchi et al., 2011). Consequently, a detailed analysis of the population structure within the training and validation sets is required before implementing genomic prediction into a breeding program.

Prediction accuracy determined in the presence of breeding populations with different performance levels would only be helpful to breeders if no prior information on population structure is available, *i.e.*, at the very beginning in breeding for a specific trait like biogas production (Riedelsheimer et al., 2012). In this case, it should be regarded whether no reduction in accuracy is found by reducing the sample size in the training set. This can be taken as an indication for the presence of hidden population structure and genotyping could be applied to identify groups of related lines. Subsequently, phenotyping a representative sample of lines from each group would be sufficient to determine differences in the performance level of the different groups (Windhausen et al., 2012a). Grouping lines according to the genetic relationship alone might not be sufficient to control for apparent population structure, because selection by breeders may result in greater differences at the phenotypic levels than reflected by genome-wide markers (Porcher et al., 2004; Pujol et al., 2008; Whitlock and Guillaume, 2009; Windhausen et al., 2012a). If population structure is apparent based on molecular, phenotypic and breeders information, the accuracy of genomic prediction should be compared to that achieved by using the mean of each breeding population in the training set. Genomic prediction will only considerably accelerate genetic gains, if prediction accuracy based on marker information is higher than that based on mean performance of breeding populations. An open question in this context is how many breeding populations should be included and how many individuals per breeding population are required to obtain high prediction accuracy.

To be effective, genomic prediction will likely require large training sets with strong relationship to the validation set (Habier et al., 2010). Furthermore, Burgueño et al. (2012) showed that for correlated environments, prediction of untested genotypes can benefit from borrowing information from correlated environments in which the training and validation sets are evaluated. These results indicate that the impact of environmental structure in combination with population structure on prediction accuracy should be considered in further research. Genomic prediction may be implemented into the CIMMYT maize breeding program to predict the performance of lines from a diversity panel, segregating lines from the same or related crosses, and progenies from closed populations within a recurrent selection program (Windhausen et al., 2012a). The breeding scenarios in which genomic prediction is most promising need to still to be defined.

Support of selection using secondary traits

The accuracy of genomic prediction is strongly dependent on the quantity and quality of phenotyping (Xu and Crouch, 2008). As costs associated to genotyping continue to decrease, providing precise estimates of the trait of interest for many genotypes is now the bottleneck in terms of maize improvement. When evaluating genotypes for grain yield and several secondary traits, it has to be regarded that breeding costs increase with each implemented trait. The costs of a secondary trait decrease with the number of genotypes that can be measured and the percentage of the genetic variance of grain yield that can be explained. The measurement of canopy and leaf reflectance explained less than 40% of the genetic variance of grain yield after validation, resulting in greater gains from direct selection (Weber et al., 2012a). Similarly, reflectance measurements have proven accurate in ranking durum wheat genotypes, although they did not provide a proper qualification of yield (Ferrio et al., 2005). Thus, selection based on predicted yield would only be suitable for pre-screening, while final yield evaluation will still be necessary. Nevertheless, genotypes the spectral reflectance pattern of which indicates low photosynthetic activity, leaf or plant water content, relative to a high yielding control could be discarded considering the reflectance spectra between 750 and 1300 nm at anthesis (Weber et al., 2012a). With selection on both sexes prior to flowering, the selection gain could be increased in comparison to selection at harvest, which can only be conducted for female plants.

Before implementing spectral reflectance measurements into a breeding program, the construction of inexpensive and easy to handle devices that can be applied to predict grain yield reliably would be desirable. Currently, the construction of a portable hyperspectral camera system is underway that can measure canopy reflectance and temperature in parallel (G. Romano, personal communication). This device should be tested in the field for its predictive ability for grain yield at anthesis across several trials, locations and years. Based on this data, a calibration model should be developed for low- and high-yielding environments separately. A combined model across test environments would not be appropriate, because this would assume that physiological mechanisms contributing to yield performance in low- and high-yielding environments are the same and more weight is given to high-yielding environments (Windhausen et al., 2012b). Nevertheless, predicted grain yield in low- and high-yielding environments may be combined by index selection as discussed above for actual grain yield.

Allocation of testing resources

Maize hybrid breeding at CIMMYT is divided into three phases: selection of lines based on *per se* performance, selection of lines based on test-cross performance, selection of hybrids in multi-location trials (Bänziger et al., 2000). During the early stages of a breeding program, numerous crosses between inbred lines and testers need to be evaluated in extensive field trials to identify hybrids with higher yield potential in the target environment. Most crosses are discarded after field evaluation due to low general performance. The best inbred lines are chosen for making single- and three-way cross hybrids, which are evaluated in optimal, managed drought and low-N environments (Bänziger et al., 2000, 2006). This strategy should be further pursued, because genotypes susceptible to low-N and drought stress at anthesis are effectively discarded while genotypes responsive to high-yielding environments are retained. Nevertheless, the heritability of managed drought stress should be increased via their conduct at locations where the soil variability is low and the staff is experienced in managing the stress through irrigation.

The best hybrids from early testing enter multi-location testing in about 60 trials. Based on nine year data, CIMMYT currently allocates its available testing resources into the conduct of 60% optimal, 22% random abiotic stress, 12% low-N and 6% managed drought trials (Weber et al., 2012b). The best 20 hybrids from the first year are reevaluated in the second year in the same composition of test environments. The resources invested into managed drought trials should be shifted to the conduct of low-N trials, because the evaluation under managed drought in multi-location trials was not predictive for performance under random abiotic stress. Investments in a larger number of low-N trials may increase selection gain, because direct selection was more efficient than indirect selection in optimal environments, genotypes can be easily selected under different levels of soil N, and performance under low-N and random abiotic stress was highly correlated. The percentage of occurrence of random abiotic stress observed in multi-location trials (22%) may not represent that in farmers' fields, because most of the CIMMYT breeding stations are currently located primarily in regions, where the probability, that 30% of the total agricultural area is affected by drought, ranges between 0 and 20% (Figure 1). Locations with high occurrence of random abiotic stress should be identified to make allocation of resources possible. For example, the number of low-yielding trials could be increased in regions where the probability of drought is higher than 30%, such as Nairobi (Kenya), Arusha (Tanzania), Bulanwayo (Zimbabwe) and El Fashir (North Sudan).

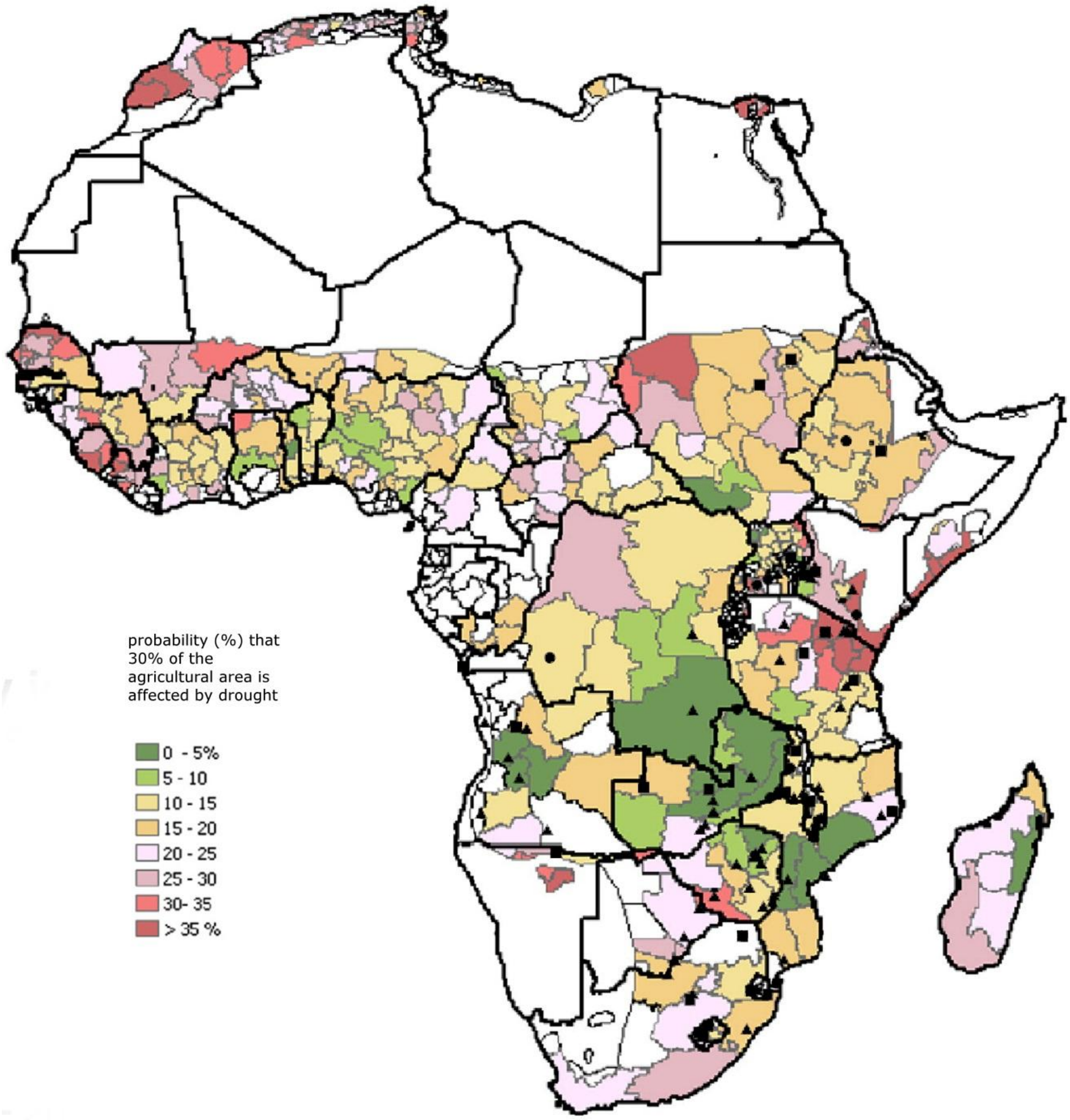

Figure 1: Probability of occurrence of having more than 30% of the agricultural area affected by drought (Rojas et al., 2011). The geographical position of breeding locations of CIMMYT and its regional partners are indicated (for further information regard Weber et al. (2012b).

The best 4-6 hybrids from multi-location testing are increased for national release testing. The evaluation on-farm completes selection on-station especially in terms of stability and responsiveness to high-yielding environments. It must be noted that if on-farm evaluation is conducted with limited replication, heritability of on-farm trials may be low compared to that achieved in on-station trials (Atlin et al., 2001).

Improvement of field designs

Often, a decrease of heritability is observed with decreasing environmental mean yield (Bänziger et al., 1997; Mandal et al., 2010). Similarly, heritability of grain yield in low-yielding environments was lower than in high-yielding environments due to a combination of lower testing effort, variable timing and intensity of random abiotic stress, and lower genetic variance and higher residual variance (Figure 2) in low-yielding environments (Weber et al., 2012b; Windhausen et al., 2012b).

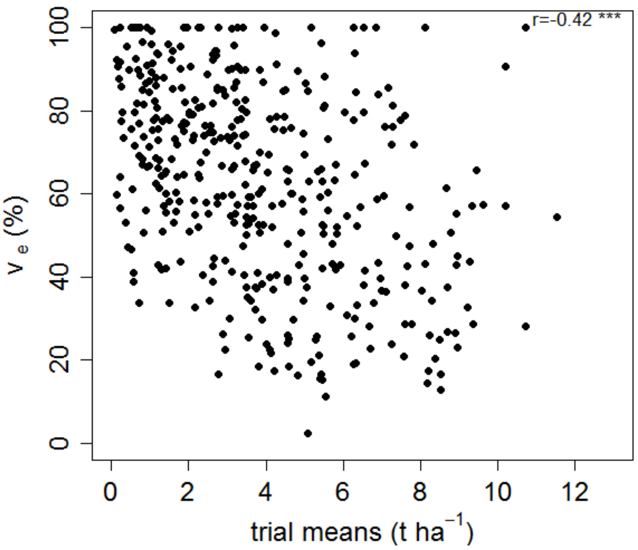

Figure 2: Relationship between trial means of grain yield and the residual variance (v_e, %) in 404 trials conducted from 2001 to 2009 for evaluating maize hybrids of early maturity. The residual variance is given as percent of the phenotypic variance calculated as the sum of the genetic and residual variance.

The adverse effect of a large residual variance in low-yielding trials could be decreased by increasing the number of field replications. Currently, multi-location trials are conducted using three field replicates (Weber et al., 2012b; Windhausen et al., 2012b). By increasing the number of replicates from three to four, the trial heritability of low-yielding environments would increase on average from 0.47 to 0.52[5]. Nevertheless, by increasing the number of field replicates while keeping the plot size constant, more space for conducting the trials is needed, which also increases the costs of testing. Alternatively, the relative importance of the residual variance can be decreased by the use of improved field designs. The use of α-lattice designs increased breeding progress in the CIMMYT

---

[5] Mean trial heritability was estimated for 170 low-yielding trials evaluating early maturing maize hybrids

maize breeding program by 20% on average relative to other lattice designs (Bänziger et al., 2000). Further gains may be achieved by using the geographic information of each plot (*i.e.*, the row and column number) for spatial analysis. Spatial modeling of trials laid out as α-lattice designs may improve gains when the spatial trend is predominantly in one direction (*e.g.*, if soil N increases in the vertical direction). In fields, where soil variability is more patchy, gains may be higher by using a row- and column design which is an extension of the α-lattice designs into two dimensions and allows to adjust for field trends in two directions (Williams et al., 2006). Analysis may proceed by fitting a baseline model with a replicate and block effect (α-lattice design) or a row and column effect (row-column design). Subsequently, it should be checked, whether addition of a spatial component is worthwhile regarding the Akaike information criterion, where the model with the lowest value is regarded as best (Müller et al., 2010; Piepho and Williams, 2010). Unfortunately, the only mixed linear model package available for *R* that can handle missing value and is open access (*lme4*) does not allow to model spatial covariance structures up to now. Alternatively, the *asreml* package[6] can be used.

In 2009, several plots per trial were allocated to in total more than 50 check varieties. Most of them were labeled as ‘local check’ without a detailed description on the pedigree or variety name. Consequently, local checks could only be used for within-trial comparison of genotypes. An attractive alternative to the use of local checks is to replicate a certain proportion of genotypes of interest for local error control, within- and across-trial comparison. Partial replication of a certain proportion of genotypes was recommended for unreplicated trials (Smith et al., 2006; Williams et al., 2011) but may also be implemented for replicated trials. For example, in the second year of multi-location testing, the 20 selected genotypes could be used for this purpose. Those genotypes could then be replicated to a greater extend across all trials, which may decrease the contribution of the residual variance and increase the heritability of low-yielding environments. Further, inclusion of long-term breeding checks would facilitate the estimation of breeding gain across time (Windhausen et al., 2012b). The selection gain of repeating selected genotypes to a higher extend as well as practical considerations for providing field designs warrant further research.

---

[6] free of charge for academic purpose
~ 600 € / year and computer for non-profit organizations
~1000€ / year and computer for commercial use

Conclusions

The results of this study showed, that the CIMMYT breeding programs in Eastern and Southern Africa should be consolidated. Selection of hybrids adapted to the target environment was ideal when combining information from optimal, low-N and random abiotic stress environments. Gains from selection were maximized by estimating hybrid means, considering the yield level effect as fixed and appropriately weighting information from all trials. To make allocation of resources possible, locations with high occurrence of random abiotic stress need to be identified. Heritability in trials conducted at these locations may be increased by decreasing the adverse effect of a large residual variance via the use of row- and column designs and/or spatial adjustment. Resources invested into managed drought trials should be maintained during early breeding stages but shifted to the conduct of low-N trials at later breeding stages. The development of a novel and inexpensive devices that facilitate grain yield prediction at anthesis on the basis of canopy reflectance and temperature may increase genetic gains and warrants further research. The prospects are promising to accelerate the breeding cycle by the use genomic prediction. Nevertheless, before incorporating genomic prediction into the CIMMYT maize breeding program, the breeding scenarios in which genomic prediction is most promising need to be defined. Generally, a detailed analysis of the population structure in the training and validation sets and the construction of larger training sets with strong genetic relationship to the validation set are recommended. Combining index selection and genomic prediction for performance in low- and high-yielding environments is a promising strategy to develop broadly-adapted maize hybrids for Eastern and Southern Africa.

References

Atlin, G.N., M. Cooper, and Å. Bjørnstad. 2001. A comparison of formal and participatory breeding approaches using selection theory. Euphytica 122(3): 463-475.

Atlin, G.N., and K.J. Frey. 1990. Selecting oat lines for yield in low-productivity environments. Crop Science 30(3): 556-561.

Braun, H.-J., G. Atlin, and T. Payne. 2010. Multi-location Testing as a Tool to Identify Plant Response to Global Climate Change. p. 115-150. *In* Reynolds, M.P. (ed.), Climate Change & Crop Production. CAB International, London, UK.

Burgueño, J., G.D.L. Campos, K. Weigel, and J. Crossa. 2012. Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. Crop Science 52: 707-719.

Bänziger, M., F.J. Betran, and H.R. Lafitte. 1997. Efficiency of high-nitrogen selection environments for improving maize for low-nitrogen target environments. Crop Science 37(4): 1103-1109.

Bänziger, M., G.O. Edmeades, D. Beck, and M. Bellon. 2000. Breeding for Drought and Nitrogen Stress Tolerance in Maize: From Theory to Practice. CIMMYT, Mexico, D.F.

Bänziger, M., P.S. Setimela, D. Hodson, B. Vivek, and M. Banziger. 2006. Breeding for improved abiotic stress tolerance in maize adapted to southern Africa. Agricultural Water Management 80(1-3): 212-224.

Ceccarelli, S., S. Grando, and J. Hamblin. 1992. Relationship between barley grain yield measured in low- and high-yielding environments. Euphytica 64(1-2): 49-58.

Ferrio, J.P., D. Villegas, J. Zarco, N. Aparicio, J.L. Araus, and C. Royo. 2005. Assessment of durum wheat yield using visible and near-infrared reflectance spectra of canopies. Field Crops Research 94(2-3): 126-148.

Habier, D., J. Tetens, F.R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genetics Selection Evolution 42(5): 1-12.

Mandal, N.P., P.K. Sinha, M. Variar, V.D. Shukla, P. Perraju, A. Mehta, A.R. Pathak, J.L. Dwivedi, S.P.S. Rathi, S. Bhandarkar, B.N. Singh, D.N. Singh, S. Panda, N.C. Mishra, Y.V. Singh, R. Pandya, M.K. Singh, R.B.S. Sanger, J.C. Bhatt, R.K. Sharma, A. Raman, A. Kumar, and G. Atlin. 2010. Implications of genotype×input interactions in breeding superior genotypes for favorable and unfavorable rainfed upland environments. Field Crops Research 118(2): 135-144.

Müller, B.U., K. Kleinknecht, J. Möhring, and H.-P. Piepho. 2010. Comparison of Spatial Models for Sugar Beet and Barley Trials. Crop Science 50(3): 794.

Piepho, H.P., and J. Möhring. 2005. Best linear unbiased prediction of cultivar effects for subdivided target regions. Crop Science 45(3): 1151-1159.

Piepho, H.P., and E.R. Williams. 2010. Linear variance models for plant breeding trials. Plant Breeding 129: 1-8.

Porcher, E., T. Giraud, I. Goldringer, and C. Lavigne. 2004. Experimental demonstration of a causal relationship between heterogeneity of selection and genetic differentiation in quantitative traits. Evolution 58(7): 1434-1445.

Przystalski, M., A. Osman, E.M. Thiemt, B. Rolland, L. Ericson, H. Østergård, L. Levy, M. Wolfe, A. Büchse, H.-P. Piepho, and P. Krajewski. 2008. Comparing the performance of cereal varieties in organic and non-organic cropping systems in different European countries. Euphytica 163(3): 417-433.

Pujol, B., A.J. Wilson, R.I.C. Ross, and J.R. Pannell. 2008. Are Q(ST)-F(ST) comparisons for natural populations meaningful? Molecular ecology 17: 4782-4785.

Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, R. Sulpice, T. Altmann, M. Stitt, L. Willmitzer, and A.E. Melchinger. 2012. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nature Genetics 44(2): 217-20.

Rojas, O., A. Vrieling, and F. Rembold. 2011. Assessing drought probability for agricultural areas in Africa with coarse resolution remote sensing imagery. Remote Sensing of Environment 115(2): 343-352.

Saatchi, M., M.C. McClure, S.D. McKay, M.M. Rolf, J. Kim, J.E. Decker, T.M. Taxis, R.H. Chapple, H.R. Ramey, S.L. Northcutt, S. Bauck, B. Woodward, J.C.M. Dekkers, R.L. Fernando, R.D. Schnabel, D.J. Garrick, and J.F. Taylor. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genetics, selection, evolution 43(40): 1-16.

Smith, A.B., P. Lim, and B.R. Cullis. 2006. The design and analysis of multi-phase plant breeding experiments. Journal of Agricultural Science 144: 393-409.

Weber, V.S., J.L. Araus, J.E. Cairns, C. Sanchez, A.E. Melchinger, and E. Orsini. 2012a. Prediction of grain yield using reflectance spectra of canopy and leaves in maize plants grown under different water regimes. Field Crops Research 128: 82-90.

Weber, V.S., A.E. Melchinger, C. Magorokosho, D. Makumbi, M. Bänziger, and G.N. Atlin. 2012b. Efficiency of managed-stress selection for maize grain yield under optimal rainfed, low-N, and random abiotic stress conditions in Southern Africa. Crop Science 52(3): 1011-1020.

Whitlock, M.C., and F. Guillaume. 2009. Testing for spatially divergent selection: comparing QST to FST. Genetics 183(3): 1055-63.

Williams, E.R., J. a John, and D. Whitaker. 2006. Construction of resolvable spatial row-column designs. Biometrics 62(1): 103-8.

Williams, E., H.-P. Piepho, and D. Whitaker. 2011. Augmented p-rep designs. Biometrical journal. Biometrische Zeitschrift 53(1): 19-27.

Windhausen, V.S., G.N. Atlin, J.M. Hickey, J. Crossa, J.-L. Jannink, M. Sorrels, B. Raman, J.E. Cairns, A. Tarekegne, K. Semagn, Y. Beyene, P. Grudloyma, F. Technow, C. Riedelsheimer, and A.E. Melchinger. 2012a. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3, Volume 2, doi: 10.1534/g3.112.003699.

Windhausen, V.S., S. Wagener, C. Magorokosho, D. Makumbi, V. Bindiganavile, H.P. Piepho, A.E. Melchinger, and G.N. Atlin. 2012b. Strategies to subdivide a target population of environments: results from the CIMMYT led maize hybrid testing programs in Africa. Crop Science 52: 2143-2152.

Xu, Y., and J.H. Crouch. 2008. Marker-assisted selection in plant breeding: From publications to practice. Crop Science 48(2): 391-407.

7. Summary

Maize is a major food crop in Africa and primarily grown by small-holder farmers under rain-fed conditions with low fertilizer input. Projections of decreasing precipitation and increasing fertilizer prices accentuate the need to provide farmers with maize varieties tolerant to random abiotic stress, especially drought and N deficiency. Genetic improvement for the target environment in Eastern and Southern Africa can be achieved by: (i) direct selection of grain yield in random abiotic stress environments, (ii) indirect selection for a secondary trait or grain yield in optimal, low-N and/or managed stress environments, or (iii) index selection using information from all test environments. At present, the maize hybrid testing programs of the International Maize and Wheat Improvement Center (CIMMYT) select primarily for grain yield under managed stress and optimal environments and subdivide the target environment according to geographic and climatic differences. It is not known to what extend the current strategy contributes to selection gains. The same holds true for genomic prediction, a strategy that is not yet implemented into the CIMMYT maize breeding program but that may accelerate breeding progress and reduce cycle length by predicting genotype performance based on molecular markers.

Regarding the different strategies mentioned for selecting high-yielding and broadly adapted maize hybrids, the breeder needs to decide which of them are most promising to increase genetic gains. Consequently, the objectives of my thesis were to (1) evaluate the potential of leaf and canopy spectral reflectance as novel secondary traits to predict grain yield across different environments, (2) estimate to what extent indirect selection in managed drought and low-N stress environments is predictive of grain yield in random abiotic stress environments, (3) investigate whether subdividing the target environment into climate, altitude, geographic, yield level or country subregions is likely to increase rates of genetic gain, and (4) evaluate the prospects of genomic prediction in the presence of population structure.

The measurement of spectral reflectance (495 – 1853 nm) of both leaves and canopy at anthesis and milk grain stage explained less than 40% of the genetic variation in grain yield after validation. Consequently, selection based on predicted grain yield is only suitable for pre-screening, while final yield evaluation will still be necessary. Nevertheless,

the prospect of developing inexpensive and easy to handle devices that can provide, at anthesis, precise estimates of final grain yield warrants further research.

Based on a retrospective analysis across 9 years, more than 600 trials and 448 maize hybrids, it was shown that maize hybrids were broadly adapted to climate, altitude, geographic and country subregions in Eastern and Southern Africa. Consequently, I recommend that the maize breeding programs of CIMMYT in the region should be consolidated. Within the consolidated breeding programs, genotypes should be selected for performance in low- and high yielding environments as the genotype-by-yield level interaction variance was high relative to the genetic variance and genetic correlations between low- and high-yielding environments were moderate. Genetic gains were maximized by index selection, considering the yield-level effect as fixed and appropriately weighting information from all trials. To allow better allocation of resources, locations with high occurrence of random abiotic stress need to be identified. Heritability in trials conducted at these locations may be increased by the use of row- and column designs and/or spatial adjustment. Furthermore, resources invested into managed drought trials should be maintained during early breeding stages but shifted to the conduct of low-N trials at later breeding stages. Investments in a larger number of low-N trials may increase selection gain, because performance under low-N and random abiotic stress was highly correlated and genotypes can be easily selected under different levels of soil N.

Prospects are promising to accelerate breeding cycles by the use of genomic prediction. Based on two large data sets on the performance of eight breeding populations, it was shown that prediction accuracy resulted primarily from differences in mean performance of these populations. Genomic prediction may be implemented into the CIMMYT maize breeding program to predict the performance of lines from a diversity panel, segregating lines from the same or related crosses, and progenies from closed populations within a recurrent selection program. The breeding scenarios in which genomic prediction is most promising still need to be defined. Generally, the construction of larger training sets with strong relationship to the validation set and a detailed analysis of the population structure within the training and validation sets are required. In conclusion, combining index and genomic selection is the most promising strategy for providing high-yielding and broadly adapted maize genotypes for the target environments in Eastern and Southern Africa.

8. Zusammenfassung

Mais ist eine der wichtigen Nahrungspflanzen in Afrika und wird vor allem von Kleinbauern ohne Bewässerung und mit limitierter Stickstoffdünung angebaut. Die Prognosen von abnehmenden Niederschlägen und steigenden Düngemittelpreisen erfordern die Züchtung von Maissorten, die eine hohe Stresstoleranz bei trockenen und stickstoffarmen Umwelten besitzen. Eine züchterische Verbesserung der Stresstoleranz kann für die Zielregionen in Ost- und Süd-Afrika durch folgende Strategien erreicht werden: (i) direkte Selektion von Kornertrag in Umwelten mit abiotischem Stress, (ii) indirekte Selektion für sekundäre Merkmale oder Kornertrag in optimalen oder kontrollierten Stressumwelten oder (iii) Index-Selektion unter Verwendung der Informationen aller Testumwelten. Derzeit selektiert das Maiszüchtungsprogramm des Internationalen Mais- und Weizenforschungszentrums (CIMMYT) in erster Linie auf Kornertrag in kontrollierten Stress- sowie optimalen Umwelten und unterteilt die Zielregion nach geographischen und klimatischen Unterschieden. Es ist nicht bekannt, inwieweit die aktuelle Strategie erfolgreich ist. Das gleiche gilt für die genomische Vorhersage anhand von genetischen Markern, einer Strategie, die im CIMMYT Maiszüchtungsprogramm künftig angewendet werden soll und den Züchtungsfortschritt erheblich beschleunigen könnte.

Hinsichtlich der erwähnten Strategien für die Selektion von hoch-ertragreichen und universal angepassten Maishybriden muss ein Züchter entscheiden, welches die vielversprechendsten sind, um den Selektionsgewinn zu erhöhen. Folglich waren die Ziele meiner Arbeit zu bewerten, inwieweit (1) sich die Messung der Lichtreflektion von Blättern und des Blätterdachs als neues sekundäres Merkmal für die Vorhersage des Kornertrags in verschiedenen Umwelten eignet, (2) indirekte Selektion in kontrollierten Stressumwelten prädiktiv ist für den Kornertrag in abiotischen Stressumwelten, (3) die Unterteilung der Zielregion anhand von Unterschieden in Klima, Höhenlage, geografischer Lage, Ertragsniveau oder Landesgrenzen den Selektionserfolg erhöht, und (4) genomische Vorhersage bei Vorliegen von Populationsstruktur in das Züchtungsprogram integriert werden kann.

Die Messung der Lichtreflektion (495 - 1853 nm) von Blättern und Blätterdach während und nach der Blüte erklärte weniger als 40% der genetischen Variation des

Kornertrags nach der Validierung. Folglich ist die Selektion anhand des vorhergesagten Kornertrags nur angemessen für eine Vorbewertung und eine Erfassung des tatsächlichen Kornertrags nachwievor notwendig. Die Konstruktion von billigen und leicht zu handhabenden Geräten, die zur Blüte eine genaue Schätzung des Kornertrags ermöglichen, rechtfertigt jedoch weitere Forschungsarbeiten.

Basierend auf einer retrospektiven Analyse über 9 Jahre, mehr als 600 Versuchen und 448 Maishybriden wurde gezeigt, dass Maishybriden adaptiert sind an verschiedene Klimata, Höhenlagen und geografische Regionen. Daher empfehle ich, dass die Zuchtprogramme von CIMMYT in Ost-und Südafrika zusammengelegt werden. Innerhalb der zusammengelegten Zuchtprogramme sollten die Genotypen für niedrig- und hoch-ertragreiche Umwelten selektiert werden, da die Interaktionsvarianz Genotyp-Ertragsniveau hoch war im Vergleich zu der genetischen Varianz und die genetischen Korrelationen zwischen niedrig- und hoch-ertragreichen Umwelten moderat waren. Der Selektionserfolg wurde durch Indexselektion maximiert, in dem das Ertragsniveau als fixer Effekt betrachtet und die Information aus allen Versuchen optimal gewichtet wurde. Um eine bessere Ressourcenallokation zu ermöglichen, sollten Standorte mit häufigem Auftreten von abiotischem Stress identifiziert werden. Die Wiederholbarkeit von Versuchen an diesen Standorten könnte durch die Verwendung von Zeilen- und Spalten-Designs und/oder räumlicher Anpassung erhöht werden. Darüber hinaus sollten die Ressourcen, die in frühen Zuchtstadien für Versuche in kontrollierten Stressumwelten investiert wurden, beibehalten werden, wohingegen sie in späteren Zuchtphasen für die Durchführung von Versuchen mit reduzierter Stickstoffdüngung verwendet werden sollten. Die Investitionen in eine größere Anzahl dieser Versuche verspricht den Zuchtfortschritt zu erhöhen, weil der Kornertrag in stickstoffarmen und abiotischen Stressumwelten hoch korreliert war und Genotypen zuverlässig unter verschiedenen Stickstoffniveaus selektiert werden können.

Die Aussichten sind vielversprechend, den Züchtungsfortschritt mit genomischer Vorhersage zu beschleunigen. Basierend auf zwei großen Datensätzen über die Leistung von acht Populationen wurde gezeigt, dass die hohe Vorhersagegenauigkeit in erster Linie auf Unterschieden in der mittleren Leistung dieser Populationen basiert. Genomische Vorhersage kann in das CIMMYT Maiszüchtungsprogramm integriert werden, um die Leistung von Linien aus einem diversem Panel, spaltenden Linien aus denselben oder verwandten Kreuzungen und Populationsnachkommen in einem rekurrentem

Selektionsprogram vorherzusagen. Die Szenarien, in denen genomische Vorhersage am vielversprechendsten ist, müssen noch genauer erforscht werden. Generell sind größere Trainingssets mit naher Verwandtschaft zum Validationsset und eine detaillierte Analyse der Populationsstruktur in den Trainings- und Validierungssets erforderlich. Die Kombination von Index- und genomischer Selektion ist die vielversprechendste Strategie, um hoch-ertragreiche und universal angepasste Maishybriden für die Zielregionen in Ost- und Südafrika bereitzustellen.

9. Acknowledgments

Foremost, I want to express my sincere thanks to Prof. Albrecht E. Melchinger for his advice, suggestions and continuous support during my work.

Sincere thanks to Gary Atlin and Elena Orsini for the many discussions and their excellent supervision! I am very indebted to Prof. H.-P. Piepho and Prof. H.-F. Utz for their advice on statistical analyses and enlightening discussions.

I had the great pleasure to collaborate with several scientists of the maize breeding program of CIMMYT in El Batan (Mexico) and Harare (Zimbabwe), the Institute of Food Crops at the Yunnan Academy of Agricultural Sciences in Mangshi and Kunming (China) and the Department of Plant Breeding and Genetics of Cornell University in Ithaca (USA). Thanks to all collaborators for sharing their knowledge with me and for their hospitality!

Many thanks to Jose-Luis Araus, Jill Cairns and Ciro Sanchez for their assistance in coordinating the field trials in Mexico, China and Thailand. Further, I owe great gratitude to all collaborators involved in conducting more than 600 trials in Eastern and Southern Africa, and Claudio Ayala and Cosmos Magorokosho for their help on constructing the data base. Thanks to in total 23, unmentioned, coauthors for contributing to my work.

I thank all my colleagues and friends at the University of Hohenheim, CIMMYT and Cornell University for the cheerful working environment. Thanks also to Silke Wagener, who substantially contributed to my work with her Master thesis. I am indebted to Christian Riedelsheimer, Jean-Luc Jannink, Jeff Endelman and John Hickey for valuable discussions and practical advices on genomic prediction. Thanks to my office mates Alexander Strigens and Matthias Martin from whom I learnt how to keep calm even at times of great disorder. I am very grateful to Heide Beck, Helga Kösling, Margit Lieb and Liliana Santamaria for resolving many organizational issues during my work.

Special thanks to Giuseppe Romano, Fan Yang, Martina Keller and Linda Homann, that accompanied me through my thesis and always brought me back to smile. My heartfelt thanks go to my family and especially my husband Kai Windhausen for encouraging me throughout my work.

I highly appreciate the financial support of this work by the BMZ 08.78603-001.00 research project.

10. Curriculum vitae

Name: Sandra VANESSA WINDHAUSEN née Weber

Date and place of birth: 12. February 1985, Marburg, Germany

Education

September 2007 – September 2009	Swiss Federal Institute of Technology, Zurich, Switzerland Master of Science ETH in Agroecosystem Science with a Major in Crop Science
October 2004 – September 2007	Justus-Liebig University Giessen, Germany Bachelor of Science, Subject Agricultural Sciences and Environmental Management
August 1995 – June 2004	Gymnasium Philippinum Weilburg, Germany High school degree (Abitur)

Agricultural internships

January 2009 – May 2012:	Two month intern and eight month doctoral fellow at the International Maize and Wheat Improvement Center (CIMMYT), El Batan, Mexico
August – November 2008	Intern at the Agroscope Reckenholz-Tänikon Research Station (ART), Zurich, Switzerland

Employment record

Since September 2012	Hybrid wheat breeder at Saaten Union France, Estrées Saint Denis, France
September 2009 – August 2012	Doctorate candidate in the institute of Plant breeding, University of Hohenheim

Sandra VANESSA WINDHAUSEN

Hohenheim, 27.07.2012

11. Erklärung

Hiermit erkläre ich an Eides statt, dass die vorliegende Arbeit von mir selbst verfasst wurde und lediglich unter Zuhilfenahme der angegebenen Quellen und Hilfsmittel angefertigt wurde. Wörtlich oder inhaltlich übernommene Stellen wurden als solche gekennzeichnet.

Die vorliegende Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Institution oder Prüfungsbehörde vorgelegt.

Insbesondere erkläre ich, dass ich nicht früher oder gleichzeitig einen Antrag auf Eröffnung eines Promotionsverfahrens unter Vorlage der hier eingereichten Dissertation gestellt habe.

____

Sandra VANESSA WINDHAUSEN

Hohenheim, July 27, 2012