Association of progeny variance and genetic distances among parents and implications for the design of elite maize breeding programs

Friedrich H Longin[1]*, Wenxin Liu[1], Nicolas Ranc[2], Jochen C Reif[1]

[1]State Plant Breeding Institute, University of Hohenheim, 70593 Stuttgart, Germany
[2]Syngenta Seeds SAS, 12, chemin de l'Hobit, B.P. 27, 31790 Saint-Sauveur, France
*Corresponding author: E-mail: friedrich.longin@uni-hohenheim.de

Abstract

Choice of crosses is crucial for a successful and sustainable management of breeding programs. Our objectives were to (1) investigate the association between the Rogers' distances among parents and the genetic variance within their crosses (σ^2_{within}) in elite maize breeding germplasm, (2) study whether this association can be improved selecting trait-specific markers, and (3) evaluate the consequences to implement the usefulness criterion based on Rogers' distances on the optimum choice of crosses. Testcross performance of eleven segregating crosses with a total of 930 progenies was evaluated in six environments for grain yield (GY) and grain moisture content (GMC). Moreover, the 930 genotypes were fingerprinted with 425 polymorphic SNP markers. Our findings revealed that working within a heterotic group, σ^2_{within} increased with increasing Rogers' distances among the parents. This was more pronounced for GY ($r_P = 0.55$ $P < 0.1$) compared to GMC ($r_P = 0.17$). Selecting trait specific markers, which were associated with putative QTL affecting these traits, led for GY to a decrease in the association between σ^2_{within} and Rogers' distances among the parents. Consequently, using for GY a regression model based on Rogers' distances estimated with an unselected set of markers allows a rough implementation of the usefulness criterion in maize breeding programs. Our model calculations suggested that implementing the usefulness criterion will facilitate a broadening of the diversity of elite maize breeding pools by counterbalancing a reduction in parental performance with an increase in σ^2_{within}.

Keywords: usefulness criterion, genetic distance, molecular markers, choice of crosses, maize

Introduction

In elite breeding programs, a large number of parental lines are available resulting in many potential crosses among them to generate new genetic variation (Longin et al, 2007). Because of limited economic resources, plant breeders have to focus on most promising crosses. One frequently used criterion for the choice of promising crosses is the average performance of the parents, which is a good predictor of the mean performance of a cross (e.g. Melchinger et al, 1998a; Utz et al, 2001; Miedaner et al, 2006). This criterion, however, neglects that the value of a particular cross depends on a linear combination of the mean of the progenies and their standard deviation (Zhong and Jannink, 2007).

Schnell and Utz (1975) suggested choosing promising crosses based on the usefulness criterion (U). U is defined as $U = \mu + \Delta G = \mu + i\, h\, \sigma_{within}$, where μ is the mean of homozygous lines that can be derived from a cross, σ_{within} is the square root of the genetic variance within this cross, i is the standardized selection intensity, and h is the square root of the heritability. While μ can be accurately predicted by the mean of the parents, robust predictors of σ^2_{within} are lacking (Zhong and Jannink, 2007). Therefore, U for cross prediction is disregarded in current breeding.

Several studies used phenotypic distances among parents to predict σ^2_{within}, but correlations have been rather low (e.g. Souza and Sorrels, 1991). Alternatively, genetic distances based on neutral molecular markers have been used for predicting σ^2_{within}. Prediction power of neutral molecular markers were also very low (e.g. Moser and Lee, 1994; Burkhamer et al, 1998; Melchinger et al, 1998a; Bohn et al, 1999). This low association can very likely be explained by (1) the small population sizes used and the corresponding large error in prediction of σ^2_{within} and (2) the use of mostly unrelated parents. For the latter, a significant correlation is only expected if substantial linkage disequilibrium exists between the quantitative trait loci contributing to σ^2_{within} and the marker loci used for estimating genetic distances (Melchinger et al, 1998a; Bohn et al, 1999). The use of unrelated parents, however, is in contrast to common praxis in elite maize breeding programs with crosses among highly related inbred lines within heterotic groups (Reif et al, 2005a).

Our objectives were to (1) investigate the association between the Rogers' distances among parents and the genetic variance within their crosses (σ^2_{within}) in elite maize breeding germplasm, (2) study whether

RECEIVED 11/05/2011

this association can be improved selecting trait-specific markers, and (3) evaluate the consequences to implement the usefulness criterion based on Rogers' distances on the optimum choice of crosses.

Materials and Methods

Association between genetic distances among parents and the genetic variance within crosses

Nine elite maize inbred lines originating from the stiff stalk heterotic group were used as parents and crossed in an incomplete diallel design (Liu et al, 2011; Table 1). Eleven segregating populations with 44 to 161 inbred progenies (in total 930) were crossed to an elite inbred tester from the non stiff stalk heterotic group and testcross progenies were evaluated in two-row plots (8.2 m²) in six environments in unreplicated trials. All plant materials used in this study are proprietary to Syngenta Seeds. Data were recorded for (1) grain yield in Mg ha^{-1}, adjusted to 155 g kg^{-1} of grain moisture (GY) and (2) grain moisture content in g kg^{-1} at harvest stage (GMC).

Genetic variances within crosses (σ^2_{within}) were estimated using PROC VARCOMP and Type I method of software package SAS (SAS Institute, 2008). Estimates of variance components due to the residuals represent confounded variance components due to genotype by environment interactions and the experimental error, because of the use of unreplicated trials.

We used Rogers' distances (RD) (Rogers, 1972) for estimation of genetic distances among parents, because it reflects the proportion of alleles that two genotypes have in common and is linearly related to the co-ancestry coefficient under certain assumptions (Reif et al, 2005b). Moreover, for testcross performance a linear relationship between RD and genotypic variance is expected assuming (1) that SNPs cover all QTL information underlying the trait under consideration, (2) no variation in the allele substitution effects across loci, and (3) absence of epistasis. RD was estimated using software Plabsoft (Maurer et al, 2008) based on 425 polymorphic SNP markers, which were evenly distributed across the genome with an average genetic map distance between adjacent markers of 2.9 cM (Liu et al, 2011).

As a second approach, we selected trait-specific markers, which were associated with GY or GMC to estimate RD among parents. We used a joint linkage association mapping approach with an additive genetic model as described by Reif et al (2010). We applied a two-step procedure for QTL detection. In a first step, stepwise multiple linear regression was used to select a set of cofactors based on the Schwarz (1978) Bayesian Criterion. Cofactor selection was performed using PROC GLMSELECT implemented in the statistical software SAS. In the second step, we calculated a P value for the association of each marker with the phenotypic value for the F test with a full model (with marker effect and cofactors) against a reduced model (with cofactors but without marker effect) (Liu et al, 2011). Genome-wide scans for QTL were conducted using statistical software R (R development core team, 2010). To select markers to estimate trait-specific RD, we applied mild selection intensity with P values smaller than 0.15 resulting for GY in 75 markers and for GMC in 102 markers.

Associations between σ^2_{within} and RD among parents were evaluated for linear and second-order polynomial relationships with software package R using multiple regression analyses following established procedures (Snedecor and Cochran, 1980). Pearson's correlation coefficients (r_p) were determined with software package R.

Model calculations to evaluate the consequences of implementing the usefulness criterion for optimum choice of crosses

Breeders tend to use best x best crosses in second cycle breeding. As a consequence, elite maize lines with high general combining ability are often highly related. Using genetically distant parents for crosses compared to elite lines is often afflicted with a reduced general combining ability. The decrease in mean performance, however, can be counterbalanced with an increased genetic variation leading also to high values for the usefulness criterion U (Schnell and Utz, 1975).

Based on our experimental data, we developed a regression model to predict the genetic variation with RD (lm(RD)) and studied the consequences of implementing the usefulness criterion for the optimum choice of crosses. We applied model calculations taking the two parents P1 and P2 with highest general combining ability in our data set as standard (UP1xP2). We then used the developed regression model between RD and σ^2_{within} to estimate the increase in RD among parents needed to counterbalance a decrease in the mean parental value by using a parent Px instead of P2 so that $U_{P1xPx} = U_{P1xP2}$. Px was estimated as $Px = P2 + 2*i*h*(lm(RD_{P1xP2})^{0.5} - lm(RD_{P1xPx})^{0.5})$. Model calculations were performed for varying RD_{P1xPx} assuming $H^2 = 0.5$ or 0.8, i for the selection of the best 25% or 5 %, P2 for GY = 13.8 Mg ha^{-1}, P2 for GMC = 220.4 g kg^{-1} and $RD_{P1xP2} = 0.25$, which represents the mean RD between parents in elite maize populations (Van Inghelandt et al, 2010).

Results and Discussion

Estimates of genetic variances within populations and heritabilities

For GY, σ^2_{within} averaged 0.23 Mg ha^{-1} with a range from 0.04 to 0.41 (Table 1). For GMC, σ^2_{within} averaged 35.65 g kg^{-1} with a minimum of 15.83 and a maximum of 76.95 g kg^{-1}. These results are in accordance to previous findings in maize (Steinhoff et al, 2011; Hallauer and Miranda 1981; Schön et al, 2004). Heritability on a plot basis ranged for GY from 0.05 to 0.25 and for GMC from 0.09 to 0.29. The phenotyp-

Table 1 - Description of the eleven segregating maize populations underlying our study, their mean, genotypic variance (σ^2_G), residual error variance (σ^2_e), and broad sense heritability on a plot basis (H^2) for grain yield (Mg ha^{-1}) and grain moisture (g kg^{-1}).

			Grain yield				Grain moisture		
Population	Progenies	Mean	σ^2_G	σ^2_e	H^2	Mean	σ^2_G	σ^2_e	H^2
PopAxD	105	13.3	0.25***	1.69	0.13	234.6	66.90***	172.8	0.28
PopAxE	161	13.5	0.22***	1.5	0.13	220.4	35.28***	172.4	0.17
PopAxF	127	13.9	0.31***	1.73	0.15	222.7	38.97***	164.7	0.19
PopAxG	93	13.5	0.41***	1.23	0.25	244.7	39.13***	123.4	0.24
PopCxE	53	13.3	0.32***	1.33	0.20	223.4	74.95***	187.4	0.29
PopDxB	92	13.7	0.08+	1.74	0.05	239.1	32.54***	177.1	0.16
PopDxH	40	13	0.19**	1.08	0.15	235.2	18.84*	167.8	0.10
PopDxI	51	12.9	0.30***	1.08	0.22	228.6	29.88**	128.1	0.19
PopExB	106	13.8	0.04	1.31	0.03	243.7	15.83**	168.7	0.09
PopExH	44	13.1	0.23**	1.28	0.15	233.8	22.17**	165.3	0.12
PopGxI	58	12.8	0.23***	0.85	0.21	228.5	17.71**	147.0	0.11
Total	930	13.5	0.29***	1.67	0.15	231.6	101.3***	242.1	0.30

adapted after Liu et al, 2011

+, *, **, *** significant at the 0.1, 0.05, 0.01 and 0.001 level of probability, respectively; heritability on an entry-mean basis was calculated as the ratio of genotypic to phenotypic variance $h^2=\sigma^2_G/(\sigma^2_G+\sigma^2_e/E)$, where E refers to the number of environments and σ^2e refers to residuals variance (confounded with σ^2_{GxE}) (Melchinger et al, 1998b)

ing of the 930 testcross progenies across six environments led to an average broad sense heritability of 0.49 for GY and 0.54 for GMC. These realized values for broad sense heritabilities represent a robust data base to investigate the association of RD and σ^2_{within}.

Association of genetic distance and σ^2_{within}

For GMC, we observed a low correlation between RD among parents and σ^2_{within} of $r_P = 0.17$, which is in line with previous findings in maize (Melchinger et al, 1998a), wheat (Utz et al, 2001), and oat (Moser and Lee, 1994). In contrast, for GY, we observed a medium correlation of $r_P = 0.55$ ($P < 0.1$) between RD among parents and σ^2_{within}. This difference can be explained by varying complexity of the traits GY and GMC as well as the use of small population sizes in the former studies resulting in large errors in estimating σ^2_{within}. Moreover, previous studies were based on mostly unrelated parents, where a significant correlation between both variables can only be expected if each DNA marker employed for estimating the RD among parents is either identical to, or else extremely tightly linked with a QTL for the trait and, vice versa, each polymorphic QTL is marked by one DNA marker (Melchinger et al, 1998a). For highly related parental lines, which is representative for elite maize breeding (Reif et al, 2005a) and used in our study, a high extent of linkage disequilibrium is expected (Stich et al, 2005). This together with our findings for GY suggests that genetic distance among parents estimated with neutral molecular markers can serve as a rough tool to implement the concept of the usefulness criterion in applied plant breeding programs.

The switch from the use of neutral markers towards selecting markers contributing to the quantitative traits might be a promising strategy to further improve the implementation of the usefulness concept in applied plant breeding programs. Therefore, we evaluated the possibility to use RD estimated based on selected trait-specific markers. For GY, the selection of markers associated with the target trait for estimation of RD led to a reduced $r_P = 0.19$ ($P < 0.5$). This finding is in accordance with results of a simulation study (Zhong and Jannink, 2007), where the use of a selected marker set compared to the full marker set led to a decrease in prediction ability of the usefulness criteria. In contrast, for GMC, selection of markers led to a slight increased $r_P = 0.30$ ($P < 0.37$). This can be explained by the lower complexity of the trait GMC compared to GY, which is in line with results from genomic selection reporting for less complex traits an increase in the prediction accuracy with approaches selecting markers instead of using all markers (Daetwyler et al, 2010). Nevertheless, more traits have to be investigated in order to draw general conclusions. Summarizing, for complex traits such as GY, genetic distances among parents should be estimated with all available neutral molecular markers to implement the usefulness criterion.

Consequences of implementation of the usefulness criterion on the optimum choice of crosses

The large number of feasible cross combinations in elite breeding programs requires the concentration on the most promising ones (Longin et al, 2007). The value of a particular cross is a linear combination of the mean of the progenies and their standard deviation (Zhong and Jannink, 2007). While the average performance of the parents is accepted as good predictor for the mean performance of a cross (e.g. Melchinger et al, 1998a; Utz et al, 2001; Miedaner et al, 2006), the prediction of the standard deviation within crosses was not yet feasible. Our findings on the positive correlation of RD among parents and σ^2_{within} for complex traits like GY suggest that implementation of the usefulness criterion is feasible.

The usefulness of a cross is maximized by taking parents with highest general combining ability and largest genetic distance among them (data not shown). Most elite lines in breeding programs, however, are highly related (e.g. Van Ingelhandt et al, 2010) and an increase in RD among parents is mostly associated with a reduced parental performance. Thus, it is of utmost interest to study to which extent

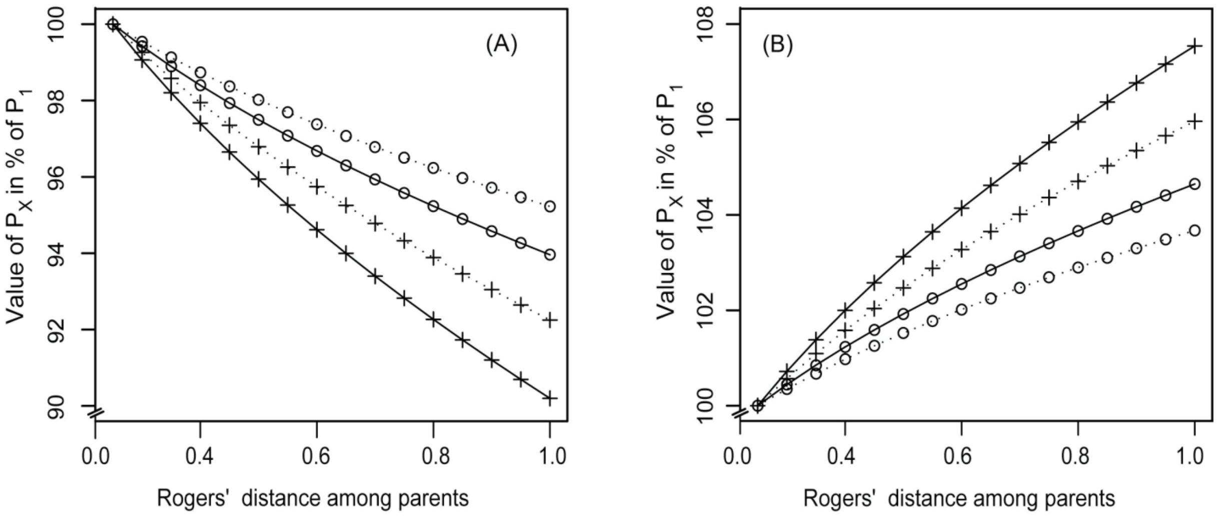

Figure 1 - Performance of parent Px required to maintain a given usefulness of a cross compared to the standard set of parents, i.e. the two lines with highest general combining ability for grain yield (A) and grain moisture content (B). (O: selected fraction = 0.25; +: selected fraction = 0.05; dotted lines: H^2=0.5; solid lines: H^2 = 0.8).

an increase in RD among parents can counterbalance a reduced performance of the parents warranting a stable value of U. For this reason, we have chosen as standard the U for the cross of the two parents with highest general combining ability and their given RD. We exchanged then parent P2 by an alternative parent Px with varying RD to parent P1 in order to determine the minimum parental performance required for Px in order to maintain the standard U defined above.

Our results suggested, that an increased RD among parents can counterbalance an important amount of reduced parental performance without affecting the usefulness of a cross (Figure 1). For instance, an increase of the RD among parents from 0.3 to 0.7 allows a reduction in parental performance of >7% and 5% for GY and GMC, respectively, while maintaining the defined standard U (Figure 1; note that an increase of GMC is negative while it is vice versa for GY). These possible reductions must be considered in a context with introduction of new variability into elite breeding populations. The best testcross performance of doubled haploid lines developed directly from landraces achieved about 87% of elite checks (Wilde et al, 2010) underlining the importance of our findings for parental selection.

The findings of our model calculations depend on the chosen values of lm(RD), RD_{P1xPx}, i and h used to determine Px. The parameters lm(RD), RD_{P1xPx}, and h depend on the underlying germplasm and are therefore specific for breeding programs and even for subpopulations within breeding programs. A validation of the developed lm(RD) model with a further experimental data set in maize (Steinhoff et al, 2011) showed little influence on Px (data not shown). In contrast, the choice of RD_{P1xP2} influences Px with an increased slope of the curves for Px with smaller RD_{P1xP2} (or vice versa for larger RD_{P1xP2}, data not shown). As RD_{P1xP2} = 0.25 represents the average RD of large commercial breeding programs (Van Ingelhandt et al, 2010), it will probably be even smaller for the very outstanding elite lines increasing the importance of σ^2_{within} for cross prediction.

Selection intensity and heritability differ strongly among breeders due to allocation of resources devoted to testing requiring more emphasis in model calculations. Selection intensity had a larger influence on Px than the heritability (Figure 1). This is due to the nature of these two variables with possible ranges for h between 0 and 1 as well as for i between 0 and values above 4. Strict selection and appropriate phenotyping commonly occur in elite breeding (cf. Schön et al, 2004) increasing i and h maximizing the importance of σ^2_{within} and the genetic distance among parents for an optimum choice of crosses. This has large implications especially for breeding schemes using genomic selection approaches, where fast series of selection cycles combined with high selection pressure rapidly leads to a reduction of the genetic diversity among elite parental lines. Consequently, research on an optimum allocation of resources with breeding programs integrating genomic selection has to consider the threat of a reduced genetic diversity causing a reduction in expected selection gain.

Acknowledgements

Thanks to helpful comments of Prof Dr HF Utz, Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim. This research was conducted within the Biometric and Bioinformatic Tools for Genomics based Plant Breeding project supported by the German Federal Ministry of Education and Research (BMBF) within the framework of GABI–FUTURE initiative. We thank two anonymous reviewers' for their suggestions to improve our manuscript.

References

Bohn M, Utz HF, Melchinger AE, 1999. Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSR and their use for predicting progeny variation. Crop Sci 39: 228-237

Burkhamer RL, Lanning SP, Martens RJ, Martin JM, Talbert LE, 1998. Predicting progeny variance from parental divergence in hard red spring wheat. Crop Sci 38: 243-248

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA, 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics 185: 1021-1031

Hallauer AR, Miranda JB, 1981. Quantitative genetics in maize breeding. 2nd edition Iowa State University Press. Ames, IA

Longin CFH, Utz HF, Melchinger AE, Reif JC, 2007. Hybrid maize breeding with doubled haploids. II. Optimum type and number of testers in two-stage selection for general combining ability. Theor Appl Genet 114: 393-402

Liu W, Gowda M, Steinhoff J, Maurer HP, Würschum T, Longin CFH, Cossic F, Reif JC, 2011. Association mapping in an elite maize breeding population. Theor Appl Genet 123: 847-858

Maurer HP, Melchinger AE, Frisch M, 2008. Population genetic simulation and data analysis with Plabsoft. Euphytica 161: 133-139

Melchinger AE, Gumber RK, Leipert RB, Vuylsteke M, Kuiper M, 1998a. Prediction of testcross means and variances among F3 progenies of F1 crosses from testcross means and genetic distances of their parents in maize. Theor Appl Genet 96: 503-512

Melchinger AE, Utz HF, Schön CC, 1998b. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and larger bias in estimates of QTL effects. Genetics 149: 383-403

Miedaner T, Schneider B, Oettler G, 2006. Means and variances for Fusarium head blight resistance of F2-derived bulks from winter triticale and winter wheat crosses. Euphytica 152: 405-411

Moser H, Lee DM, 1994. RFLP variation and genealogical distance, multivariate distance, heterosis, and genetic variance in oats. Theor Appl Genet 87: 947-956

R Development Core Team, 2010. R: A language and environment for statistical computing, R Foundation for Statistical Computing. http://www.R-project.org

Reif JC, Hamrit S, Heckenberger M, Schipprack W, Maurer HP, Bohn M, Melchinger AE, 2005a. Trends in genetic diversity among European maize cultivars and their parental components during the past 50 years. Theor Appl Genet 111: 838-845

Reif JC, Melchinger AE, Frisch M, 2005b. Genetical and mathematical properties of similarity coefficients applied in plant breeding and seed bank management. Crop Sci 45: 1-7

Reif JC, Liu W, Gowda M, Maurer HP, Möhring J, Fischer S, Schechert A, Würschum T, 2010. Genetic basis of agronomically important traits in sugar beet (*Beta vulgaris* L) investigated with joint linkage association mapping. Theor Appl Genet 8: 1489-1499

Rogers JS, 1972. Measures of genetic similarity and genetic distance, pp 145-153. In: Studies in genetics VII. Publ 7213. University of Texas, Austin

SAS Institute, 2008. SAS/STAT user's guide, version 9.2. SAS Institute, Cary, NC

Schnell FW, Utz HF, 1975. F1-Leistung und Elternwahl in der Züchtung von Selbstbefruchtern, pp 243-248. In: Bericht über die Arbeitstagung der Vereinigung Österreichischer Pflanzenzüchter. BAL Gumpenstein, Gumpenstein

Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE, 2004. Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167: 485-498

Schwarz G, 1978. Estimating the dimension of a model. Ann Stat 6: 461-464

Snedecor GW, Cochran WG, 1980. Statistical methods. Iowa State University Press. Ames

Souza E, Sorrels ME, 1991. Prediction of progeny variation in oat from parental genetic relationships. Theor Appl Genet 82: 233-241

Steinhoff J, Liu W, Maurer HP, Würschum T, Longin CFH, Ranc N, Reif JC, 2011. Multiple-line cross QTL-mapping in European elite maize. Crop Sci 51: 2505-2516

Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC, 2005. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor Appl Genet 111: 723-730

Utz HF, Bohn M, Melchinger AE, 2001. Predicting progeny means and variances of winter what crosses from phenotypic values of their parents. Crop Sci 41: 1470-1478

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B, 2010. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120: 1289-1299

Wilde K, Burger H, Prigge V, Presterl T, Schmidt W, Ouzunova M, Geiger HH, 2010. Testcross performance of doubled-haploid lines developed from European flint maize landraces. Plant Breed 129: 181-185

Zhong S, Jannink JJ, 2007. Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. Genetics 177: 567-576